

JOURNAL OF CONSULTING PSYCHOLOGY

Edited By
Edward S. Bordin
University of Michigan

Associate Editor
Boyd McCandless
University of Iowa

Advisory Editors

John W. Atkinson
University of Michigan

Frank X. Barron
University of California, Berkeley

John M. Butler
University of Chicago

Allen T. Dittmann
National Institute of Mental Health

Seymour Fisher
Baylor University

Harrison G. Gough
University of California, Berkeley

William L. Hays
University of Michigan

Robert B. Malmö
McGill University

Paul E. Meehl
University of Minnesota

Daniel R. Miller
University of Michigan

Ruth L. Munroe
New York, New York

Alan K. Rosenwald
University of Illinois

Saul Rosenzweig
Washington University

Julian B. Rotter
Ohio State University

Nevitt Sanford
University of California, Berkeley

Seymour B. Sarason
Yale University

Robert R. Sears
Stanford University

Edwin S. Shneidman
Veterans Administration, Los Angeles, California

Read D. Tuddenham
University of California, Berkeley

VOLUME 24
1960



Arthur C. Hoffman, *Managing Editor*; Helen Orr, *Promotion Manager*

Published Bimonthly By
The American Psychological Association, Inc.
1333 Sixteenth St., N.W.
Washington 6, D. C.

Printed at
Lancaster, Pennsylvania

Copyright © 1960 by the American Psychological Association, Inc.

Contents of Volume 24

<i>Bandura, Albert, Lipsher, David, and Miller, Paula E.</i> Psychotherapists' Approach-Avoidance Reactions to Patients, Expressions of Hostility - - - - -	1
<i>Barrows, Gordon A., and Zuckerman, Marvin.</i> Construct Validity of Three Masculinity-Femininity Tests - - - - -	441
<i>Bassin, Alexander.</i> See <i>Smith, Alexander B.</i>	
<i>Beck, Karl W.</i> See <i>Nichols, Robert C.</i>	
<i>Becker, Wesley C.</i> The Relationship of Factors in Parental Ratings of Self and Each Other to the Behavior of Kindergarten Children as Rated by Mothers, Fathers, and Teachers	507
<i>Bendig, A. W.</i> Age Differences in the Interscale Factor Structure of the Guilford-Zimmerman Temperament Survey - - - - -	134
<i>Bendig, A. W.</i> Factor Analyses of "Anxiety" and "Neuroticism" Inventories - - - - -	161
<i>Bendig, A. W.</i> The Factorial Validity of Items on the IPAT Anxiety Scale - - - - -	374
<i>Blau, Theodore H., and Schaffer, Robert E.</i> The Spiral Aftereffect Test (SAET) as a Predictor of Normal and Abnormal Electroencephalographic Records in Children - - - - -	35
<i>Block, Jack.</i> See <i>Chang, Judy.</i>	
<i>Bolduc, Thomas E.</i> Social Value-Need Patterns in Mental Retardates - - - - -	472
<i>Briggs, Peter F., and Wirt, Robert D.</i> Intra-Q Deck Relationships as Influences and Realities in Personality Assessment - - - - -	61
<i>Broen, William E., Jr.</i> Ambiguity and Discriminating Power in Personality Inventories -	174
<i>Buss, Arnold H.</i> See <i>Ferguson, Donald C.</i>	
<i>Caine, T. M.</i> The Expression of Hostility and Guilt in Melancholic and Paranoid Women	18
<i>Canter, Arthur.</i> The Efficacy of a Short Form of the MMPI to Evaluate Depression and Morale Loss - - - - -	14
<i>Cartwright, Rosalind Dymond, and Vogel, John L.,</i> A Comparison of Changes in Psychoneurotic Patients During Matched Periods of Therapy and No Therapy - - - - -	121
<i>Cattell, Raymond B., and McMichael, Robert E.</i> Clinical Diagnosis by the IPAT Music Preference Test - - - - -	333
<i>Chance, June Elizabeth.</i> Personality Differences and Level of Aspiration - - - - -	111
<i>Chang, Judy, and Block, Jack.</i> A Study of Identification in Male Homosexuals - - - - -	307
<i>Chapman, Loren J.</i> See <i>Gottesman, Leonard.</i>	
<i>Clayton, William H.</i> See <i>Ellsworth, Robert B.</i>	
<i>Cleveland, Sidney E.</i> Body Image Changes Associated with Personality Reorganization -	256
<i>Cochran, Irene.</i> See <i>Iscoe, Ira.</i>	
<i>Consalvi, Conrad.</i> See <i>Quay, Herbert C.</i>	
<i>Cooper, Ruth.</i> Objective Measures of Perception in Schizophrenics and Normals - - - -	209
<i>Corah, Norman L.</i> See <i>Feldman, Marvin J.</i>	
<i>Corotto, Loren V., and Curnutt, Robert H.</i> The Effectiveness of the Bender-Gesalt in Differentiating a Flight Group from an Aggressive Group of Adolescents - - - - -	368
<i>Crowne, Douglas P., and Marlowe, David.</i> A New Scale of Social Desirability Independent of Psychopathology - - - - -	349
<i>Curnutt, Robert H.</i> See <i>Corotto, Loren V.</i>	
<i>Dahlstrom, W. Grant.</i> See <i>Meehl, Paul E.</i>	
<i>Davitz, Joel R.</i> Manifest Anxiety and Social Behavior - - - - -	556
<i>Davitz, Joel R.</i> See <i>Schonbar, Rosalea A.</i>	
<i>Davitz, Joel R., and Mason, Donald J.</i> Manifest Anxiety and Social Perception - - - -	554
<i>Day, R. H.</i> The Aftereffect of Seen Movement and Brain Damage - - - - -	311
<i>Dean, Sanford J.</i> The Generality of Expectancy Statements as a Function of Situational Definition - - - - -	558

<i>Dennis, Wayne, and Raskin, Evelyn.</i> Further Evidence Concerning the Effect of Hand-writing Habits upon the Location of Drawings - - - - -	548
<i>Dinoff, Michael.</i> Subject Awareness of Examiner Influence in a Testing Situation - - -	465
<i>Dobbins, D. A., Stockwell, F. E., and Loving, W. S.</i> Individual and Social Correlates of Prison Escapes - - - - -	95
<i>Doidge, William T., and Holtzman, Wayne H.</i> Implications of Homosexuality among Air Force Trainees - - - - -	9
<i>Donohoe, John W.</i> A Dimensional Analysis of Clinical Judgement - - - - -	96
<i>Duker, Jan.</i> See <i>Gilberstadt, Harold.</i>	
<i>Efron, Herman Y.</i> An Attempt to Employ a Sentence Completion Test for the Detection of Psychiatric Patients with Suicidal Ideas - - - - -	156
<i>Ellsworth, Robert B., and Clayton, William H.</i> The Effects of Chemotherapy on Length of Stay and Rate of Return for Psychiatrically Hospitalized Patients - - - - -	50
<i>Fager, Robert E.</i> Relation of Rorschach Movement and Color Responses to Cognitive Inhibition - - - - -	276
<i>Feldman, Marvin J., and Corah, Norman L.</i> Social Desirability and the Forced Choice Method - - - - -	480
<i>Ferguson, Donald C., and Buss, Arnold H.</i> Operant Conditioning of Hostile Verbs in Relation to Experimenter and Subject Characteristics - - - - -	324
<i>Fiske, Donald W., Howard, Kenneth, and Rechenberg, William.</i> The EPPS Profile Stability Coefficient - - - - -	370
<i>Freedman, Laurence Zelic.</i> See <i>Kaswan, J.</i>	
<i>Freides, David.</i> Toward the Elimination of the Concept of Normality - - - - -	128
<i>Froehlich, Abraham.</i> See <i>Smith, Alexander, B.</i>	
<i>Gilberstadt, Harold, and Duker, Jan.</i> Case History Correlates of Three MMPI Profile Types	361
<i>Ginott, Haim G.</i> A Rationale for Selecting Toys in Play Therapy - - - - -	243
<i>Glucksberg, Sam.</i> See <i>Katz, Irwin.</i>	
<i>Goldberg, Philip A.</i> On the Breakdown of the Sense of Reality: A Comment - - - - -	415
<i>Goldman, Rosaline.</i> Changes in Rorschach Performance and Clinical Improvement in Schizophrenia - - - - -	403
<i>Gottesman, Leonard, and Chapman, Loren J.</i> Syllogistic Reasoning Errors in Schizophrenia	250
<i>Gottlieb, Ann Lodge, and Parsons, Oscar A.</i> A Coaction Compass Evaluation of Rorschach Determinants in Brain Damaged Individuals - - - - -	54
<i>Gough, Harrison G.</i> Theory and Measurement of Socialization - - - - -	23
<i>Grater, Harry.</i> Impulse Repression and Emotional Adjustment - - - - -	144
<i>Griffith, Richard M., and Taylor, Vivian H.</i> Incidence of Bender-Gestalt Figure Rotations	189
<i>Grigg, Austin E., and Thorpe, Joseph S.</i> Deviant Responses in College Adjustment Clients: A Test of Berg's Deviation Hypothesis - - - - -	92
<i>Grosz, Hanus J.</i> See <i>Levitt, Eugene E.</i>	
<i>Gruen, Walter.</i> Rejection of False Information about Oneself as an Indication of Ego Identity - - - - -	231
<i>Grunebaum, Margaret B.</i> See <i>Wallach, Michael A.</i>	
<i>Gurrslin, Orville.</i> See <i>Hunt, Raymond G.</i>	
<i>Hare, A. Paul, Waxler, Nancy, Saslow, George, and Matarazzo, Joseph D.</i> Simultaneous Recording of Bales and Chapple Interaction Measures During Initial Psychiatric Interviews - - - - -	193
<i>Hartup, Willard W., and Zook, Elsie A.</i> Sex-Role Preferences in Three- and Four-Year-Old Children - - - - -	420
<i>Heilbrun, Alfred B., Jr.</i> Perception of Maternal Child Rearing Attitudes in Schizophrenics	169
<i>Heilizer, Fred.</i> An Exploration of the Relationship between Hypnotizability and Anxiety and/or Neuroticism - - - - -	432
<i>Holtzman, Wayne H.</i> See <i>Doidge, William T.</i>	

<i>Honigfeld, Gilbert, and Spigel, Irwin M.</i> Achievement Motivation and Field Independence	550✓
<i>Hood-Williams, J.</i> The Results of Psychotherapy with Children: A Revaluation - - -	84
<i>Howard, Kenneth.</i> See <i>Fiske, Donald W.</i>	
<i>Hozier, Ann.</i> A Reply to Goldberg - - - - -	417
<i>Hunt, Raymond G., Roach, Jack L., and Gurrslin, Orville.</i> Social-Psychological Factors and the Psychiatric Complaints of Disturbed Children - - - - -	194✓
<i>Iscoe, Ira, and Cochran, Irene.</i> Some Correlates of Manifest Anxiety in Children - - -	97✓
<i>Izard, Carroll E.</i> Personality Characteristics Associated with Resistance to Change - -	437
<i>Jenkin, Noël, and Morse, Sally A.</i> Size-Distance Judgment in Organic Mental Defectives	139
<i>Judson, Abe J., and MacCasland, Barbara W.</i> The Effects of Chlorpromazine on Psychological Test Scores - - - - -	192
<i>Kagan, Jerome.</i> The Long Term Stability of Selected Rorschach Responses - - - - -	67
<i>Kanfer, Frederick H., Phillips, Jeanne S., Matarazzo, Joseph D., and Saslow, George.</i> Experimental Modification of Interviewer Content in Standardized Interviews - - -	528
<i>Kaswan, J., Wasman, M., and Freedman, Lawrence Zelic.</i> Agression and the Picture-Frustration Study - - - - -	446
<i>Katz, Irwin, Glucksberg, Sam, and Krauss, Robert.</i> Need Satisfaction and Edwards PPS Scores in Married Couples - - - - -	205
<i>Kemp, David E.</i> See <i>Parsons, Oscar A.</i>	
<i>King, Gerald F., and Schiller, Marvin.</i> Ego Strength and Type of Defensive Behavior - -	215
<i>Kingsley, Leonard.</i> Wechsler-Bellevue Patterns of Psychopaths - - - - -	373
<i>Klaber, M. Michael.</i> Manifestations of Hostility in Neurodermatitis - - - - -	116
<i>Klein, Donald C.</i> Some Concepts Concerning the Mental Health of the Individual - - -	288✓
<i>Kleinmuntz, Benjamin.</i> An Extension of the Construct Validity of the Ego Strength Scale	463 -
<i>Knowles, John B.</i> The Temporal Stability of MPI Scores in Normal and Psychiatric Populations - - - - -	278
<i>Korman, Maurice.</i> Ego Strength and Conflict Discrimination: An Experimental Construct Validation of the Ego Strength Scale - - - - -	294 -
<i>Korman, Maurice.</i> Implicit Personality Theories of Clinicians as Defined by Semantic Structures - - - - -	180✓
<i>Krauss, Robert.</i> See <i>Katz, Irwin.</i>	
<i>L'Abate, Luciano.</i> Personality Correlates of Manifest Anxiety in Children - - - - -	342✓
<i>Lachmann, Frank M.</i> Perceptual-Motor Development in Children Retarded in Reading Ability - - - - -	427
<i>Levitt, Eugene E.</i> Reply to Hood-Williams - - - - -	89
<i>Levitt, Eugene E., and Grosz, Hanus J.</i> A Comparison of Quantifiable Rorschach Anxiety Indicators in Hypnotically Induced Anxiety and Normal States - - - - -	31
<i>Lingoes, James C.</i> MMPI Factors of the Harris and the Wiener Subscales - - - - -	74
<i>Lipsher, David.</i> See <i>Bandura, Albert.</i>	
<i>Liverant, Shephard.</i> Intelligence: A Concept in Need of Re-examination - - - - -	101
<i>Lorr, Maurice.</i> See <i>McNair, Douglas M.</i>	
<i>Lothrop, William W.</i> Psychological Test Covariates of Conceptual Deficit in Schizophrenia	496
<i>Loving, W. S.</i> See <i>Dobbins, D. A.</i>	
<i>Luckey, Eleanore Braun.</i> Martial Satisfaction and Parent Concepts - - - - -	195✓
<i>MacCasland, Barbara W.</i> See <i>Judson, Abe J.</i>	
<i>McDonough, Joseph M.</i> Critical Flicker Frequency and the Spiral Aftereffect with Process and Reactive Schizophrenics - - - - -	150
<i>Machover, S.</i> See <i>Plumeau, F.</i>	
<i>McMichael, Robert E.</i> See <i>Cattell, Raymond B.</i>	
<i>McNair, Douglas M., and Lorr, Maurice.</i> Therapists' Judgements of Appropriateness of Psychotherapy Frequency Schedules - - - - -	500
<i>Marlowe, David.</i> See <i>Crowne, Douglas P.</i>	

<i>Marshall, Simone.</i> Personality Correlates of Peptic Ulcer Patients - - - - -	218
<i>Martire, John G.</i> See <i>Palermo, David S.</i>	
<i>Mason, Donald J.</i> See <i>Davitz, Joel R.</i>	
<i>Matarazzo, Joseph D.</i> See <i>Hare, A. Paul.</i>	
<i>Matarazzo, Joseph D.</i> See <i>Kanfer, Frederick H.</i>	
<i>Mattsson, Patrik D.</i> Communicated Anxiety in a Two-Person Situation - - - - -	488
<i>Meehl, Paul E., and Dahlstrom, W. Grant.</i> Objective Configural Rules for Discriminating Psychotic from Neurotic MMPI Profiles - - - - -	375
<i>Messick, Samuel.</i> Dimensions of Social Desirability - - - - -	279
<i>Miller, Paula E.</i> See <i>Bandura, Albert.</i>	
<i>Mitchell, James V., Jr., and Pierce-Jones, John.</i> A Factor Analysis of Gough's California Psychological Inventory - - - - -	453
<i>Morse, Sally A.</i> See <i>Jenkin, Noël.</i>	
<i>Murstein, Bernard I.</i> Factor Analyses of the Rorschach - - - - -	262
<i>Neel, Ann Filinger.</i> Inhibition and Perception of Movement on the Rorschach - - - - -	224
<i>Nichols, Robert C., and Beck, Karl W.</i> Factors in Psychotherapy Change - - - - -	388
<i>Ogdon, Donald P.</i> WISC IQs for the Mentally Retarded - - - - -	187
<i>Palermo, David S., and Martire, John G.</i> The Influence of Order of Administration on Self-Concept Measures - - - - -	372
<i>Parsons, Oscar A., and Kemp, David E.</i> Intellectual Functioning in Temporal Lobe Epilepsy	408
<i>Parsons, Oscar A.</i> See <i>Gottlieb, Ann Lodge.</i>	
<i>Peterson, Donald R.</i> See <i>Quay, Herbert C.</i>	
<i>Phillips, Jeanne S.</i> See <i>Kanfer, Frederick H.</i>	
<i>Pierce-Jones, John.</i> See <i>Mitchell, James V., Jr.</i>	
<i>Plumeau, F., Machover, S., and Puzzo, F.</i> Wechsler-Bellevue Performances of Remitted and Unremitted Alcoholics, and Their Normal Controls - - - - -	240
Psychological Test Reviews - - - - -	99, 466
<i>Puzzo, F.</i> See <i>Plumeau, F.</i>	
<i>Quay, Herbert C., Peterson, Donald R., and Consalvi, Conrad.</i> The Interpretation of Three Personality Factors in Juvenile Delinquency - - - - -	555
<i>Raskin, Evelyn.</i> See <i>Dennis, Wayne.</i>	
<i>Rechenberg, William.</i> See <i>Fiske, Donald W.</i>	
<i>Roach, Jack L.</i> See <i>Hunt, Raymond G.</i>	
<i>Rubin, Mandel, and Shontz, Franklin C.</i> Diagnostic Prototypes and Diagnostic Processes of Clinical Psychologists - - - - -	234
<i>Santorum, Aldo.</i> A Cross-Validation of the House-Tree-Person Drawing Indices Predicting Hospital Discharge of Tuberculosis Patients - - - - -	400
<i>Saslow, George.</i> See <i>Hare, A. Paul.</i>	
<i>Saslow, George.</i> See <i>Kanfer, Frederick H.</i>	
<i>Schaffer, Robert E.</i> See <i>Blau, Theodore H.</i>	
<i>Schein, Jerome D.</i> The Duration of the Archimedes Spiral Afterimage in the Diagnosis of Brain Damage - - - - -	299
<i>Schiller, Marvin.</i> See <i>King, Gerald F.</i>	
<i>Schonbar, Rosalea A., and Davitz, Joel R.</i> The Connotative Meaning of Sexual Symbols - -	483
<i>Schroeder, Pearl.</i> Client Acceptance of Responsibility and Difficulty of Therapy - - -	467
<i>Seidel, Claudene.</i> The Relationship between Klopfer's Rorschach Prognostic Rating Scale and Phillips' Case History Prognostic Rating Scale - - - - -	46
<i>Shontz, Franklin C.</i> See <i>Rubin, Mandel.</i>	
<i>Silverstein, A. B.</i> A Cluster Analysis of Object Sorting Behavior - - - - -	98
<i>Simkins, Lawrence.</i> Examiner Reinforcement and Situational Variables in a Projective Testing Situation - - - - -	541

<i>Smith, Alexander B., Bassin, Alexander, and Froehlich, Abraham.</i> Change in Attitudes and Degree of Verbal Participation in Group Therapy with Adult Offenders - - - - -	247
<i>Smith, Philip A.</i> A Factor Analytic Study of the Self-Concept - - - - -	191
<i>Sperber, Zanwil.</i> An Empirical Evaluation of a Test of Role-Playing Ability - - - - -	557
<i>Spigel, Irwin M.</i> See <i>Honigfeld, Gilbert.</i>	
<i>Stockwell, F. E.</i> See <i>Dobbins, D. A.</i>	
<i>Storms, Lowell H.</i> Rationales for the "Twisted Pear" - - - - -	552
<i>Stricker, George.</i> See <i>Zax, Melvin.</i>	
<i>Strupp, Hans H.</i> See <i>Wallach, Martin S.</i>	
<i>Tallarico, Robert B.</i> See <i>Williams, Carl D.</i>	
<i>Taylor, Vivian H.</i> See <i>Griffith, Richard M.</i>	
<i>Tedeschi, James T.</i> See <i>Williams, Carl D.</i>	
<i>Thorpe, Joseph S.</i> See <i>Grigg, Austin E.</i>	
<i>Ulrich, David N.</i> See <i>Wallach, Michael A.</i>	
<i>Vogel, John L.</i> See <i>Cartwright, Rosalind Dymond.</i>	
<i>Wagner, Nathaniel N.</i> Development Aspects of Impulse Control - - - - -	537
<i>Wallach, Martin S., and Strupp, Hans H.</i> Psychotherapists' Clinical Judgments and Attitudes towards Patients - - - - -	316
<i>Wallach, Michael A., Ulrich, David N., and Grunebaum, Margaret B.</i> Relationship of Family Disturbance to Cognitive Difficulties in a Learning-Problem Child - - - - -	355 ✓
<i>Waller, Patricia F.</i> A Comparison of Shading Responses Obtained with Two Rorschach Methodologies from Psychiatric and Nonpsychiatric Subjects - - - - -	43 ✓
<i>Wasman, M.</i> See <i>Kaswan, J.</i>	
<i>Waxler, Nancy.</i> See <i>Hare, A. Paul.</i>	
<i>Weinberg, Jon R.</i> A Further Investigation of Body-Cathexis and the Self - - - - -	277
<i>Williams, Carl D., Tallarico, Robert B., and Tedeschi, James T.</i> Manifest Needs and Manifest Anxiety - - - - -	371
<i>Wirt, Robert D.</i> See <i>Briggs, Peter F.</i>	
<i>Zax, Melvin, and Stricker, George.</i> The Effect of a Structured Inquiry on Rorschach Scores - - - - -	328
<i>Zook, Elsie A.</i> See <i>Hartup, Willard W.</i>	
<i>Zuckerman, Marvin.</i> The Development of an Affect Adjective Check List for the Measurement of Anxiety - - - - -	457 ✓
<i>Zuckerman, Marvin.</i> See <i>Barrows, Gordon A.</i>	

PSYCHOTHERAPISTS' APPROACH-AVOIDANCE REACTIONS TO PATIENTS' EXPRESSIONS OF HOSTILITY

ALBERT BANDURA, DAVID H. LIPSHER, AND PAULA E. MILLER

Stanford University

In a previous study (Bandura, 1956), an inverse relationship was found between therapists' anxiety level and ratings of their psychotherapeutic competence. The purpose of the present investigation was to determine, through a response by response analysis of actual patient-therapist interactions, the specific ways in which therapists' anxieties may affect their psychotherapeutic work. Since hostility conflicts tend to be present to some degree in all patients, it was decided to single out for study the patient-therapist interaction around the expression of hostility.

Although it may be possible to specify the conditions essential for effecting certain changes in a patient's behavior, whether or not a particular therapist can produce these conditions may be partly determined by his own personality characteristics. A minimal condition for the resolution of a patient's conflicts would seem to be that the patient's conflictive feelings are permitted to occur within the therapy situation. If, however, certain classes of expression are anxiety provoking for the therapist, he is less likely to permit or encourage the patient to express himself in these ways. Moreover, when the patient does express tendencies that are threatening to the therapist, the anxieties he elicits often motivate the therapist to avoid a continuation of the anxiety producing interaction (Cutler, 1958; Little, 1951; Luborsky, 1952; Reich, 1951; Rigler, 1956). In line with this reasoning, it was predicted that therapists who display high hostility anxiety would be more likely to respond to patients' hostility with avoidance reactions and less likely to respond with approach reactions than would therapists who show low hostility anxiety.

Hypotheses were also advanced concerning the effect of the therapists' approach and avoidance reactions on the patients' behavior. It was assumed that the therapists' approach reactions would constitute positive reinforcements for the patients and that such reactions would therefore serve to encourage the patients to express feelings of hostility. It was further assumed that the therapists' avoidance reactions would serve either as nonrewards or as negative reinforcements and would have the effect of decreasing or inhibiting the expression of hostile feelings. It seemed possible in some instances, particularly where the patients are strongly instigated to hostility, that the therapists' negative responses would lead the patients to displace the hostility to some other object rather than to discontinue all expressions of hostility. It was therefore predicted that if the patients were to continue in the hostility topic following the therapists' avoidance responses they would change the object of their hostility more often than they would following therapists' approach responses.

METHOD

Subjects. The tape recorded interviews of 17 parents who were undergoing psychotherapy at a parent-child clinic provided the data for this study. Twelve different therapists, 2 females and 10 males, were represented. All were advanced clinical psychology students, and all except 2 were undergoing individual psychotherapy at the time of the study.

Therapists' personality characteristics. The therapists' personality characteristics were assessed through ratings provided by four clinical psychology staff members who had extensive contact, both socially and professionally, with the student therapists. Each therapist was rated independently by three staff members on eight 5-point rating scales.

Hostility anxiety, which was central to this study,

was assessed in three ways. In the first place, it was assumed that with high anxiety there would be a greater amount of indirect than of direct hostility so that, if hostility were expressed at all, its forms would more likely be indirect. Ratings were consequently made of the amount of *direct* and of the amount of *indirect hostility* that the therapists typically displayed in their interactions with others. The mere absence of hostility, whether of a direct or indirect form, would not in itself necessarily indicate the presence of anxiety. For instance, one might expect a person to show little hostility if he has experienced little frustration or instigation to hostility. On the other hand, if a person strongly inhibits any expression of hostility in the face of strong instigation, the presence of anxiety would seem to be indicated. Therefore, ratings were also made of the degree to which the therapists inhibited any expression of hostility. The *hostility inhibition* scale is given below.

This scale deals with the *readiness* with which the S gets irritated, annoyed, angered when frustrated, provoked or thwarted, e.g., when delayed, inconvenienced, criticized, opposed, etc.

1. Very easily irritated and angered—expresses irritation, annoyance, anger at the slightest frustration
2. Easily irritated and angered—expresses irritation, annoyance, anger even when mildly frustrated, provoked, or thwarted
3. Expresses irritation, annoyance, anger when moderately frustrated, provoked, or thwarted
4. Difficult to arouse to anger or to irritate—expresses irritation, annoyance, anger only when strongly frustrated, provoked, or thwarted
5. Very difficult to arouse to anger or to irritate—practically never expresses irritation, annoyance, or anger when strongly frustrated, provoked, or thwarted.

The presence of hostility anxiety was thus inferred from the therapist's tendency to inhibit hostile reactions in the face of high instigation and from his tendency to respond with predominantly indirect forms of hostility.

In addition to the three measures of hostility, the therapists were also rated on several scales measuring dependency behavior: amount of *help seeking*, amount of *approval seeking*, amount of *dependency inhibition*; they were also rated on a scale measuring the extent of *sex inhibition*; and on a scale measuring *warmth*. These latter measures were not used as a basis for any predictions.

Procedure for rating patient-therapist interaction. From the total number of interviews available for the 17 patients, 110 interviews (approximately two thirds of the total) were randomly selected for analysis. Coding of these interviews was done by two judges who worked independently on different interviews. However, in order to provide an estimate of the reliability of the judges' ratings, a sample of 20 interviews, randomly selected from the pool of 110, was independently coded by both judges. In the case

of these 20 interviews, the mean coding was used in the further analyses.

In order to insure independence of the coding, the judges had no knowledge of the ratings of the therapists' personality characteristics.

Scoring unit. The unit scored was an *interaction sequence* beginning with a patient statement, the therapist's response, and the immediately following patient response. An example is given below:

P.: I had more work to do than I felt I should do, that I had to do, work which I felt was unimportant, wasn't necessary to be done. I was told to do it. I didn't think it was right or necessary at all.

T.: How do you feel when you have to do something he wants you to do?

P.: Well, I don't like it but I do it.

Patient response category. The patient category scored was hostility which was defined as any expression of dislike, resentment, anger, antagonism, opposition, or of critical attitudes. In order to study changes in the object toward whom the patient directed hostility, this category was further divided into six subcategories according to the referent: (a) spouse, (b) children, (c) parents, (d) self, (e) therapist, and (f) other persons or objects.

Therapist response categories. The therapists' responses were divided into two general classes: approach and avoidance reactions.

Approach reactions included verbal responses that were primarily designed to elicit from the patient further expressions of hostile feelings, attitudes, and behavior. The following classes of responses were included in the approach category:

a. *Approval.* Therapist sanctions and expresses explicit agreement with the patient's hostile feelings or behavior.

P.: I don't know but I got so mad when he came home.

T.: Under the circumstances, how could you have felt otherwise?

b. *Exploration.* Therapist asks for further clarification, elaboration, and detailing of the patient's hostile feelings or behavior.

P.: For some reason I had a bad day, just couldn't fall asleep. I just felt aggravated.

T.: Aggravated? Can you tell me a little about that?

c. *Instigation.* Therapist shifts the discussion from a nonhostility topic to the hostility topic or reintroduces the hostility expressed by the patient earlier in the interview or in previous interviews.

P.: I was pretty fatigued when I got home. I went to bed right off the bat but I just tossed and turned. When I woke up in the morning I had this pain in the leg. I went to the doctor and he gave me a shot.

T.: Yes. Let's get back again to that evening, that irritated feeling you had.

d. *Reflection*. Therapist repeats or restates the patient's hostile feelings.

P.: And when the kids don't listen to me it rubs me the wrong way. I lose my temper.

T.: You get mad.

e. *Labeling*. Therapist names the patient's feelings, attitudes or behavior as hostile; points out patterns in the patient's feelings or behavior; suggests relationships between present hostile feelings and behavior and past experiences.

P.: I've picked up the tabs on that girl all my life. Always paying, always costing money. I think I resented her a good deal.

T.: Maybe some of your resentment for Joyce is partly displaced from your mother in that your mother shipped her to you to take care of. You were mad at your mother as well as Joyce.

Avoidance reactions included those verbal responses designed to inhibit, discourage, or divert the patients' hostile expressions. The following classes of responses were included in the avoidance category:

a. *Disapproval*. Therapist is critical of the patient's having felt or behaved in a hostile, aggressive manner.

P.: So I blew my top and hit her.

T.: Just for that you hit her?

b. *Topical transition*. Therapist changes the discussion from the hostility topic to a nonhostility topic.

P.: My mother annoys me.

T.: How old is your mother?

c. *Silence*. Therapist makes no verbal response for 4 sec. or more after the patient has expressed hostility.

P.: I just dislike it at home so much at times.

T.: [Silence]

P.: So I just don't know what to do.

d. *Ignoring*. Therapist responds to the content of the patient's response but ignores the hostile affect.

P.: I lose my temper over his tardiness.

T.: What are the results of his being tardy?

e. *Mislabeling*. Therapist labels as nonhostile feelings that are clearly hostile.

P.: When are you going to give me the results of those tests? I think I'm entitled to know.

T.: You seem to be almost afraid to find out.

Unclassified responses included unscorable utterances and responses that were irrelevant to the above subcategories.

In coding the interviews, the record was played and each patient statement was judged as either hostile or nonhostile. If a statement was rated as hostile, both the therapist response and the immediately succeeding patient response were rated. The object of the patient's hostility was also noted.

It would have been desirable to have used separate sets of judges for rating the therapist and patient responses respectively. This could have been ac-

TABLE 1
INTERJUDGE AGREEMENT IN SCORING RESPONSE UNITS

	Number of Units
Perfect agreement	261
Single discrepancy	
Therapist's response	72
Hostility following the therapist's response	16
Object of hostility	12
Two discrepancies	
Therapist's response and object	5
Therapist's response and hostility	7
Object and hostility	1
Two discrepancies on object of hostility	2
Unit scored by one judge only	81

complished only through the use of typescripts from which therapist or patient statements had been deleted (Dittman, 1952). On the other hand, direct recordings had certain advantages. In many instances a patient's hostility is communicated more through voice cues than through content cues and had the ratings been made from typescripts some of the more subtle hostile responses would have been lost. As a partial control, however, the judge rated the therapist response before rating the immediate response it elicited in the patient.

The scores yielded by the content analysis were the number of interaction sequences in which therapists responded with approach or avoidance reactions to the patients' hostility and the number of sequences in which hostility followed therapists' approach and avoidance, respectively. The hourly totals were then summed for each therapy case.

RESULTS

Interjudge Reliability

Since the unit of analysis is an interaction sequence, the most relevant index of reliability is the degree of agreement between the judges in coding the response units. There are a number of ways in which discrepancies can occur. The judges may disagree as to whether or not the patient is expressing hostility and/or disagree as to the object toward whom the hostility is directed. They may differ in scoring the therapist's response. Finally, they may disagree on whether or not hostility is expressed in the immediately following patient response and/or in the object of the patient's hostility.

TABLE 2
FREQUENCY OF SPECIFIC INTERJUDGE DISCREPANCIES

	Number of Units
Therapists' responses	
Reflection—labeling	13
Reflection—exploration	11
Reflection—Approval	2
Exploration—Instigation	7
Exploration—Labeling	4
Labeling—Approval	2
Ignoring—Topical transition	4
Ignoring—Mislabeling	3
Mislabeling—Topical transition	1
Ignoring—Exploration	22
Ignoring—Labeling	5
Ignoring—Reflection	9
Exploration—Topical transition	1
Object of hostility	
General—self	4
General—therapist	3
General—spouse	4
General—child	1
Therapist—self	3
Therapist—spouse	3
Therapist—child	1
Spouse—child	3

The degree of interjudge agreement in coding the response units was obtained from the 20 interviews that the two judges rated in common. Comparison of the ratings revealed that in 261 of the units scored the judges were in perfect agreement. In 100 of the units they showed only minor discrepancies, most of which involved differences in their rating of the therapists' responses (Table 1).

The frequency with which specific types of scoring disagreements occurred are presented in Table 2. Many of the discrepancies in coding the therapists' responses involved categories which overlap to some degree as, for example, reflection and labeling which are responses on an interpretative continuum. More marked discrepancies occurred, however, in 37 units where one judge rated the therapists' responses as avoidance whereas another judge rated them as some form of approach reaction.

The reliability of the ratings of the thera-

pists' personality characteristics was estimated by the use of an analysis of variance technique developed by Ebel (1951). The results are given in Table 3.

Therapists' Characteristics and the Use of Approach-Avoidance Responses

Of the 4734 interaction sequences that were scored, 1619—approximately 34%—were ones in which the patients expressed hostility. The therapists, in turn, responded with 938 approach and 859 avoidance reactions. In order to test whether differences in therapists' personality characteristics were related to the relative frequency of their approach of avoidance interventions, the therapists were dichotomized into a high scoring and low scoring group on each of the personality scales. The measure of the therapists' relative preference for approach or avoidance reactions was based on the ratio of the number of approach to the number of avoidance responses. The significance of the differences between the two groups was then tested by means of the Mann-Whitney *U* test (Siegel, 1956).

Of the three hostility anxiety measures, only one—direct hostility—yielded a statistically significant difference. Therapists who expressed their hostility in direct forms were more likely to respond with approach reactions when the patients expressed hostility toward extratherapeutic objects than were therapists who were rated low on the direct hostility scale. This difference was significant at the .03 level. These two groups of therapists did not differ, however, in their handling of hostility when the therapist himself was the object of the patient's hostility.

TABLE 3
RELIABILITIES OF THE RATINGS OF THE
THERAPISTS' CHARACTERISTICS

Scale	<i>r</i>
Direct hostility	.86
Indirect hostility	.71
Hostility inhibition	.80
Help seeking	.72
Approval seeking	.72
Dependency inhibition	.69
Sex inhibition	.72
Warmth	.86

TABLE 4
PERCENTAGE OF APPROACH REACTIONS TO
PATIENTS' HOSTILITY

Therapist	Object of Hostility	
	Therapist	Others
A	50%	85%
B	— ^a	75
C	50	65
D	50	58
E	43	56
F	— ^a	53
G	— ^a	52
H	49	51
I	30	50
J	38	43
K	20	44
L	— ^a	35

^a The interviews that were coded in these cases yielded no responses in the hostility toward therapist category.

Only one of the remaining personality measures was related to the therapists' handling of hostility. Therapists who displayed a high need for approval were more likely to avoid the patients' hostility, whether directed toward the therapist ($p < .06$) or toward others ($p < .002$), than were therapists who were rated low on the approval seeking scale.

In order to determine which of the specific approach and avoidance reactions may have contributed to the over-all differences that were obtained, separate statistical analyses were made for each of the therapist subcategories. The findings showed that therapists who readily expressed hostility in direct forms were less likely to ignore the patients' hostility than were therapists who were rated low on the direct hostility scale ($p < .03$). In addition, therapists high in approval seeking were less likely to make efforts to explore the patients' hostility ($p < .08$), were more likely to ignore hostility ($p < .004$), and were more inclined to change the discussion from hostility to nonhostility ($p < .02$) than were therapists who displayed low approval seeking.

In general, therapists were less likely to approach the patients' hostility when the therapist was the object of hostility than when the patient directed his hostility toward others (Table 4). This difference, tested by means of the sign test for the difference between two

correlated sets of scores (Siegel, 1956), was significant at the .004 level.

Influence of Therapists' Approach and Avoidance Reactions on Patients' Expression of Hostility

The influence of the therapist's behavior on the patient was measured in terms of the frequency with which the patient continued to express hostility following a particular therapist intervention. These results are presented in Table 5. Very few outright disapproval responses were scored, consequently, this category was not included in the analysis.

As can be seen from Table 5, if a therapist approached the patient's hostility, the patient was almost certain to continue expressing such feelings. On the other hand, if the therapist responded with an avoidance reaction, the patient was likely to drop the hostility topic. In testing for the significance of these differences, each specific approach category was compared with the total avoidance category. Similarly, each specific avoidance category was compared with total approach. The significance levels, yielded by the sign test, are given in Table 6.

Although the different approach reactions appeared to be equally effective in eliciting hostile feelings, some variability was noted among the avoidance reactions, with silence and mislabeling drawing hostility significantly more often than the remainder of the avoid-

TABLE 5
PERCENTAGE OF TIMES HOSTILITY FOLLOWED
SPECIFIC THERAPIST RESPONSES

Therapist Response	Total Hostility
Total approach	92%
Reflection	93
Labeling	92
Exploration	90
Instigation	93
Approval	84
Total avoidance	43
Ignoring	41
Topical transition	22
Silence	60
Mislabeling	70

TABLE 6

SIGNIFICANCE OF DIFFERENCES BETWEEN THE
APPROACH AND AVOIDANCE CATEGORIES

Categories Compared		Total Hostility <i>p</i>
Total approach	vs. total avoidance	.002
Reflection	vs. total avoidance	.002
Labeling	vs. total avoidance	.002
Exploration	vs. total avoidance	.003
Instigation	vs. total avoidance	.002
Approval	vs. total avoidance	.01
Ignoring	vs. total approach	.002
Topical transition	vs. total approach	.001
Silence	vs. total approach	.02
Mislabeling	vs. total approach	.01
Topical transition	vs. remainder of avoidance	.02
Silence	vs. remainder of avoidance	.05
Mislabeling	vs. remainder of avoidance	.02

ance interventions. Inasmuch as silence is a rather ambiguous cue, the patient may at times interpret the therapist's behavior as indicative of interest and attentiveness rather than a sign of avoidance, disinterest, or disapproval. In such instances, the therapist's silence might be responded to by the patient with further expressions of hostility. As for mislabeling, such responses oftentimes led the patient to express his feelings even more directly so as to correct the apparent failure in communication.

Table 7 presents the findings on changes in

TABLE 7

PERCENTAGE OF TIMES PATIENTS CHANGED THE OBJECT
OF HOSTILITY FOLLOWING SPECIFIC
THERAPISTS' RESPONSES

Therapist Response	% of Object Changes
Total approach	32%
Reflection	30
Labeling	42
Exploration	29
Approval	35
Total avoidance	70
Ignoring	72
Topical transition	81
Silence	59
Mislabeling	52

the object of hostility following the therapists' responses.

In contrast to approach reactions, the therapists' avoidance reactions were far more likely to lead the patients to change the object toward whom they were expressing their hostility. These differences, tested by means of the sign test, were clearly significant for the over-all categories as well as for the subcategories (Table 8).

Again, we find that silence and mislabeling had less of a disruptive effect on the patients' expression of hostility than did the other

TABLE 8

SIGNIFICANCE LEVELS OF DIFFERENCES BETWEEN
APPROACH AND AVOIDANCE CATEGORIES

Categories Compared		<i>p</i>
Total approach	vs. total avoidance	.01
Reflection	vs. total avoidance	.01
Labeling	vs. total avoidance	.01
Exploration	vs. total avoidance	.01
Approval	vs. total avoidance	.01
Ignoring	vs. total approach	.01
Topical transition	vs. total approach	.02
Silence	vs. total approach	.02
Mislabeling	vs. total approach	.02
Silence	vs. remainder of avoidance	.05
Mislabeling	vs. remainder of avoidance	.05

avoidance responses. This finding is in accordance with that reported in the preceding section (Table 5).

DISCUSSION

The hypotheses concerning the influence of therapists' hostility anxieties on their handling of hostility in the psychotherapeutic interaction were only partially confirmed. While therapists who expressed hostility in direct forms tended to encourage their patients to express hostility, no significant differences were found between therapists who differed in the amount of indirect hostility or in the degree of hostility inhibition that they displayed. The failure to find significant differences for the latter measures may be due, in part, to the relative homogeneity of the therapists under study. Few, if any, of them mani-

fested severe hostility conflicts; and few, if any, were completely free of hostility anxieties.

It was of interest to find that therapists who displayed a high need for approval tended to discourage the expression of hostility on the part of their patients. One might assume that a therapist who is strongly motivated to elicit and maintain the approval of others would generalize to some degree the approval seeking to his patient as well. Thus, the therapist may refrain from encouraging the patient to express feelings that may be somewhat anxiety or guilt provoking, since to make the patient uncomfortable may lose him the patient's approval.

During the course of psychotherapy the patient may generalize to the therapist some of his conflicts and emotional reactions. Once these responses are transferred, their modification will depend on the therapist's response. Since the therapist has control over most of the conditions that may be essential for relearning to occur, the patient's behavior can be more readily modified in relationship to the therapist than in extratherapeutic interactions where interpersonal consequences are apt to be somewhat inconsistent or even contradictory. Our findings suggest, at least as far as the handling of hostility is concerned, that therapists are less inclined to accept and to analyze hostile feelings directed toward themselves than they are to accept and analyze similar feelings expressed toward others. It is difficult to know how much this finding may be a function of the relative inexperience of the therapists in the study, the majority of whom were still in training.

Most previous studies that have investigated the influence of interviewers' behavior on patients' productions have focused on positive reinforcing responses (Krasner, 1958; Rogers, 1958) and, except for Murray's (1956) investigation, there is little data on the effect of therapists' negative responses on patients' verbalizations. From the results of this study it is clear that avoidance reactions occurred almost as frequently as did those of approach. The avoidance reactions, however, tended to take the form of nonrewards rather than of frankly negative reinforcements. Disapproval was very rarely encountered, whereas

same form of nonresponse was much more common.

The prediction that approach responses would be more effective than avoidance responses in getting the patient to express his hostility feelings was clearly confirmed. Not only were the patients apt to drop the discussion of hostility if the therapist responded with avoidance, but also, in instances where they did continue with the topic, they were more likely to change the object of their hostility. In evaluating these results, it is important to bear in mind that the criterion of the influence of the therapists' behavior was the immediately observable effect it had on the patients' verbal behavior. Had some criterion of delayed outcome been used, the results may very well have been different. Moreover, the study was limited to the patients' verbal behavior and the accompanying verbal reactions of the therapists. The operation of reward and punishment, on the other hand, is apt to take more subtle and nonverbal forms, particularly in the case of disapproval where there is little sanction for such reactions in the psychotherapeutic relationship. It is highly probable that many nonverbal cues, which in themselves could have produced the changes in the patients' behavior, accompanied the therapists' verbal responses.

A number of the personality measures did not appear to be related to the therapists' handling of hostility. They may, however, be related to the therapists' handling of other forms of behavior; e.g., sexual and dependency expressions. A study of such relationships is now being planned.

SUMMARY

The present investigation was designed to test the following hypotheses: (a) that therapists with high hostility anxiety are more inclined to avoid rather than to approach patients' hostility than are therapists who display low anxiety; (b) that approach reactions encourage the patient to express further hostility, whereas avoidance reactions serve to decrease or inhibit such expression; and (c) that patients are more likely to change the object of their hostility following an avoidance response by the therapist than they are following an approach response.

To test these hypotheses, tape recordings of 110 interviews obtained from 17 patients treated by 12 therapists were analyzed. The therapists' personality characteristics were assessed through ratings provided by clinical psychology staff members. Patient-therapist interaction sequences were coded for the number of times therapists responded with approach or avoidance reactions to the patients' hostility and also for the frequency with which patients continued to express hostility immediately following therapists' approach or avoidance interventions. The objects toward whom the patients directed their hostility were also rated.

The results obtained were as follows:

1. Therapists who typically expressed their own hostility in direct forms and who displayed low need for approval were more likely to permit and encourage their patients' hostility than were therapists who expressed little direct hostility and who showed high approval seeking behavior. No significant relationships were found for six other measures of the therapists' characteristics.

2. In general, psychotherapists were more inclined to avoid hostility when it was directed toward themselves than when the patients directed their hostility toward other objects.

3. The patients were more likely to drop the hostility topic or to change the object of

their hostility following therapists' avoidance reactions than they were following the therapists' approach reactions.

REFERENCES

- BANDURA, A. Psychotherapists' anxiety level, self-insight, and psychotherapeutic competence. *J. abnorm. soc. Psychol.*, 1956, 52, 333-337.
- CUTLER, R. L. Countertransference effects in psychotherapy. *J. consult. Psychol.*, 1958, 22, 349-356.
- DITTMAN, A. T. The interpersonal process in psychotherapy: Development of a research method. *J. abnorm. soc. Psychol.*, 1952, 47, 236-244.
- EBEL, R. L. Estimation of reliability of ratings. *Psychometrika*, 1951, 16, 507-524.
- KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, 55, 148-170.
- LITTLE, M. Counter-transference and the patient's response to it. *Int. J. Psycho-Anal.*, 1951, 32, 32-40.
- LUBORSKY, L. The personality of the psychotherapist. *Menninger Quart.*, 1952, 6, 1-6.
- MURRAY, E. J. The content-analysis method of studying psychotherapy. *Psycho. Monogr.*, 1956, 70(13, Whole No. 420).
- REICH, ANNIE. On counter-transference. *Int. J. Psycho-Anal.*, 1951, 32, 25-31.
- RIGLER, D. Some determinants of therapist behavior. Unpublished doctoral dissertation, Univer. of Michigan, 1956.
- ROGERS, J. M. Operant conditioning in a quasi-therapy situation. Unpublished doctoral dissertation, Stanford Univer., 1958.
- SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

(Received February 11, 1959)

IMPLICATIONS OF HOMOSEXUALITY AMONG AIR FORCE TRAINEES

WILLIAM T. DOIDGE

USAF Hospital, Lackland Air Force Base

AND WAYNE H. HOLTZMAN

University of Texas

During the past decade, research on sexual behavior has rapidly gained wider acceptance by the general public, and many commonly held attitudes are seriously challenged by new information. One major issue concerns the nature of homosexuality. Is homosexuality symptomatic of a more general personality disturbance or may it be regarded, in some instances at least, as confined to the sexual sector alone? In a recent article Hooker (1957) suggests, albeit very cautiously:

Even if one assumes that homosexuality represents a severe form of maladjustment to society in the sexual sector of behavior, this does not necessarily mean the homosexual must be severely maladjusted in other sectors of his behavior. Or if one assumes that homosexuality is a form of severe maladjustment internally, it may be that the disturbance is limited to the sexual sector alone (p. 30).

A contrasting opinion is advanced by the Group for the Advancement of Psychiatry (1955, p. 2): "When such homosexual behavior persists in an adult, it is then a symptom of severe emotional disorder." It is with this conflicting opinion that the present paper is concerned.

As part of a more extensive project, recent studies conducted at Lackland Air Force Base were analyzed to see which, if either, of the two views could be supported. From a military point of view the issue is one of practical importance in deciding upon proper disposition of homosexual cases brought to light in the Air Force.

The GAP report quite properly distinguishes between homosexuality and homosexual behavior. Homosexuality is considered to be a "persistent emotional and physical at-

traction to members of the same sex," and according to this report "constitutes an abnormal personality development." Adolescent exploratory experiences, occasional contacts while intoxicated, or isolated acts are more appropriately called homosexual behavior as distinguished from the condition of homosexuality. This distinction is especially important in the military service, where homosexual activity may result in trial by courtmartial and separation from the Air Force. Occasions also arise when it is necessary to identify those individuals who pretend to be homosexuals for the purpose of obtaining a discharge.

There are a number of unique advantages to a study of male homosexuals in the Air Force. Nearly 100,000 airmen a year pass through basic training at Lackland Air Force Base under controlled conditions, providing a large sample of young adult men representing all characteristics of the general population. Most previous research on homosexuality has been severely limited by strong biases in the population studied. The almost complete lack of privacy and the barracks-type of living are likely to stimulate sexual drives in male homosexuals. Prevailing cultural attitudes and stringent military policy heighten the conflict, precipitating contact with the psychiatric clinic.

In the Air Force all homosexual suspects are subjected to exhaustive interrogations by special investigators. Since heterosexual subjects (Ss) are not usually put through this procedure, it was deemed advisable in formulating an experimental design to include a heterosexual group which had been through such an investigation. In the final formula-

Table 1

Categories of Experimental Groups According to Homosexuality-Heterosexuality
Dimension and Reason for Special Investigation

Group	N	Psychosexual Orientation	Reason for Investigation
H Homosexual (predominant)	20	Predominantly homosexual	Homosexual behavior
A Homosexual (accessory)	20	Predominantly heterosexual with homosexual experience	Homosexual behavior
D Heterosexual (disciplinary)	20	Exclusively heterosexual	Nonsexual offenses
N Heterosexual (normal)	20	Exclusively heterosexual	Not investigated

tion two homosexual and two heterosexual groups, each containing 20 individuals, were selected for thorough case study and psychological testing.

Description of the Study

All homosexual suspects referred to the psychiatric clinic at Lackland Air Force Base during a period of several months were given an experimental battery of psychological tests, a searching psychiatric interview, and a thorough case history. The psychiatrist¹ conducting the interview rated each individual on a six-point heterosexual-homosexual scale adapted from Kinsey's classification (Kinsey, Pomeroy, & Martin, 1948). The scale ranged from exclusively heterosexual orientation, through various mixtures of heterosexuality and homosexuality, to exclusively homosexual behavior.

Twenty airmen in each of four groups were selected for the present study. The 20 men judged to be exclusively homosexual comprised the H group. The A group consisted of men who were predominantly heterosexual but had experienced varying degrees of homosexual contact. Since individuals of both homosexual groups had been subjected to special investigation as suspected homosexuals, a control group (D) was formed, consisting of heterosexuals who had been through a similar investigation for other kinds of alleged

offenses. A fourth group (N) of normal heterosexual men constituted a second control group for comparison with the two classes of homosexuals.

All 80 men were given a battery of 10 tests in the following order: Sexual Identification Survey, Homosexual Homonyms, Edwards Personal Preference Schedule (Edwards, 1955), Heineman's Forced-Choice Anxiety Scale (Heineman, 1953), Worchel's Self-Activities Inventory (Worchel, 1957), the Food Preference and Aversion Scale, the Rorschach test (including a special testing of limits for sex), the Blacky Pictures Technique (Blum, 1950), the MMPI, and six subtests of the Wechsler Adult Intelligence Scale (Vocabulary, Information, Similarities, Picture Completion, Block Design, and Digit Symbol). The Sexual Identification Survey was designed especially for this study and consisted of 30 human figures taken from drawings, paintings, statues, and other sources which were photographically reproduced in such a way as to make the sex ambiguous. The S's task was to identify the sex of each figure. The Food Preference and Aversion Scale was adapted from Wallen (1943) and consisted of 29 foods to be rated by the S on a five-point like-dislike scale.

The third test developed especially for this study was the Homosexual Homonyms. Using Secord's (1953) method, a list of 10 homonyms (blow, gay, fairy, French, fruit, queer, rear, suck, pansy, Greek) was embedded in a buffer list of 50 words and given to each S as a word association test.

¹ All psychiatric examinations were accomplished by Louis J. West, Department of Psychiatry and Neurology, University of Oklahoma School of Medicine, Oklahoma City, Oklahoma.

Each test was scored without awareness of the sexual class into which the *S* had been placed in order to avoid any possible bias. Where score distributions were reasonably normal, comparisons among the four experimental groups were made by analysis of variance. Where distributions were severely skewed or truncated, chi square or the Kruskal-Wallis nonparametric test (Walker & Lev, 1953) was used.

Results

Striking differences were obtained between the H group and the other three groups on most of the MMPI scales. *F* ratios significant beyond the .01 level were obtained for *F*, *Hy*, *Pd*, *Mf*, *Pa*, *Pt*, and *Sc*; and beyond the .05 level for *Hs*, *D*, and *Si*. Mean *T* scores for each scale and each group are presented in Fig. 1.

The sharp difference in elevation of most scores in the H group, as compared to the other three groups, strongly suggests that the symptom of homosexuality is but a part of a more general personality disturbance. The close similarity of the A group to the two control groups is convincing evidence of the

important distinction between the exclusively homosexual individual and the person who has only occasionally engaged in homosexual activities. Of particular interest is the comparison of the four groups on the masculinity-femininity scale. Only two men in the H group had *T* scores on the *Mf* scale less than 70 (and one of these obtained a score of 67)! By contrast, the highest *T* score obtained in the A group was 70—and there were only 4 such individuals out of 20. None of the normal controls (N) and only one of the delinquent controls (D) had an *Mf* score as high as 70.

Similar results were obtained for Heine- man's Forced-Choice Anxiety Scale. Adapted from Taylor's Manifest Anxiety Scale (Taylor, 1953) and the MMPI, Heineman's scale was designed to eliminate social desirability as a response set in anxiety questionnaires. All of the statistical significance in the *F* ratio between groups can be attributed to the unusually high mean score for the H group. The H, A, D, and N groups obtained mean scores of 46.6, 35.4, 30.5, and 31.5 respectively. Again, it is apparent that the exclusively homosexual male manifests a general

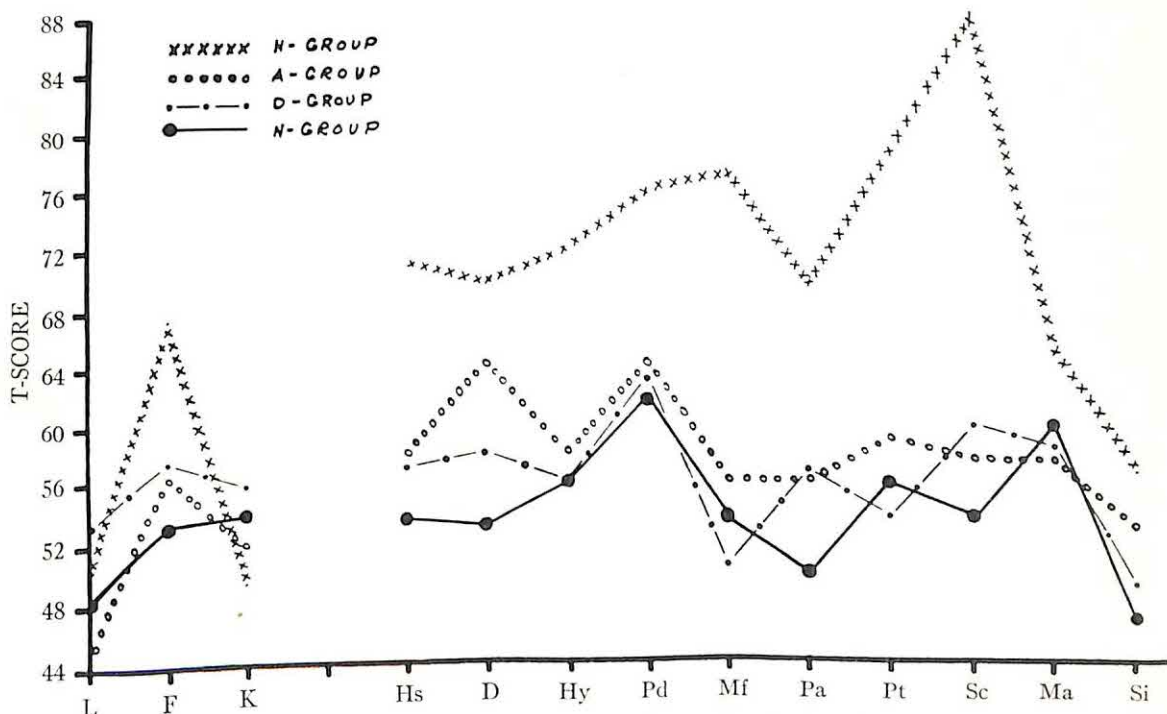


Fig. 1. Mean *T* scores of the MMPI for four groups of airmen (predominantly homosexual, accessory homosexual, disciplinary, and normal). *N* = 80; each group = 20.

personality disturbance while the occasional homosexual does not.

Assuming that sexual associations to the critical words of the Homosexual Homonyms are anxiety arousing, delayed response times may reflect this anxiety. The mean transformed scores for response times yield identical rankings among the groups with the scores obtained from Heineman's Forced-Choice Anxiety Scale, and the magnitudes of the differences are almost identical, suggesting that the similarity is due to the effects of anxiety.

Food aversions have been positively correlated with anxiety and neuroticism (Smith, Powell, & Ross, 1955; Wallen, 1945). Combining the scores of the two homosexual groups and the two heterosexual groups on the Food Preference and Aversion Scale provided adequate frequencies for contingency tables. The total list and 5 of the 29 foods yielded aversion scores for the combined homosexual groups significantly different from the combined heterosexual groups beyond the .05 level.

Based upon the assumption that sexual confusion may exist in a person of one sex who has the interests, attitudes, and motivations of the other, the Sexual Identification Survey was administered to all Ss. While response times and deviation scores were generally consistent with a priori expectations, none of the differences were statistically significant. These negative findings suggest that homosexuals, although possibly confused in their masculine-feminine roles, may have learned to compensate for their confusion. In order to seek out others who are also homosexual, they may have become perceptually attuned to recognize subtle cues of another's homosexuality. This training may compensate for any role confusion in themselves or in their identification with others. In any event, these results suggest that popular interpretations of masculine-feminine role confusion in projective techniques may be invalid.

Perhaps the most important implication to emerge from this study concerns the severity of psychopathology which is likely to accompany the markedly homosexual individual. For this person, the findings support the contention that he is likely to be suffering from

an emotional disorder which is relatively pervasive and severe. By comparison, the accessory homosexual does not show evidence of extensive personality disturbance of the severity found to characterize the markedly homosexual individual.

Another implication is the need for developing assessment instruments which will increase the accuracy and the objectivity of the evaluation of a homosexual condition with respect to both its presence and its severity. Such information would have diagnostic, prognostic, and treatment values. Some of the tests used in this study show developmental potentialities for such purposes.

Further research similar in kind and scope to the present study should be undertaken for the female sex. The same kinds of data for homosexual women could be integrated with data from this study to provide more complete information on homosexuality in general. Complementary findings from both studies might contribute to a better understanding of psychosexual development, over-all personality functioning, and the relationship of homosexuality to psychopathology.

Summary

A battery of 10 psychological tests, including 2 original tests, was administered to 80 enlisted airmen divided equally into four groups. Ss were assigned to groups according to whether they were heterosexual, partly homosexual, or markedly homosexual in their psychosexual orientations. Only the markedly homosexual group gave test records that were strikingly different from the control groups, suggesting that markedly homosexual individuals are likely to be suffering from an emotional disorder which is relatively pervasive, severe, and disqualifying for military service. The partly homosexual group (composed of individuals who were predominantly heterosexual but with varying degrees of homosexual experience) gave test records that closely approximated the results of the two control groups. Implications for improving existing methods of assessment and disposition of homosexual cases are discussed, and extension of the present study to include female members of the Air Force is recommended.

REFERENCES

- BLUM, G. S. *The Blacky Pictures. Manual.* New York: Psychological Corp., 1950.
- EDWARDS, A. L. *Edwards Personal Preference Schedule. Manual.* New York: Psychological Corp., 1955.
- GROUP FOR THE ADVANCEMENT OF PSYCHIATRY, Committee on Cooperation with Governmental (Federal) Agencies. Report on homosexuality with particular emphasis on this problem in governmental agencies. (Rep. No. 30.) New York, N. Y.: Author, 1955. Pp. 1-7.
- HOOKE, EVELYN. The adjustment of the male overt homosexual. *J. proj. Tech.*, 1957, 21, 1-31.
- HEINEMAN, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, 17, 447-454.
- KINSEY, A. C., POMEROY, W. B., & MARTIN, C. E. *Sexual behavior in the human male.* Philadelphia: Saunders, 1948.
- SECORD, P. F. Objectification of word-association procedures by the use of homonyms: A measure of body cathexis. *J. Pers.*, 1953, 21, 479-495.
- SMITH, W., POWELL, E. K., & ROSS, S. Manifest anxiety and food aversion. *J. abnorm. soc. Psychol.*, 1955, 50, 101-104.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- WALKER, H. M., & LEV, J. *Statistical inference.* New York: Holt, 1953.
- WALLEN, R. Sex differences in food aversions. *J. appl. Psychol.*, 1943, 27, 288-298.
- WALLEN, R. Food aversions of normal and neurotic males. *J. abnorm. soc. Psychol.*, 1945, 40, 77-81.
- WEST, L. J., DOIDGE, W. T., & WILLIAMS, R. L. An approach to the problem of homosexuality in the military service. *Amer. J. Psychiat.*, 1958, 115, 392-401.
- WORCHEL, P. Adaptability screening of flying personnel: Development of a Self-Concept Inventory for predicting maladjustment. *USAF Sch. Aviat. Med. Rep.*, 1957, No. 56-62.

(Received December 23, 1958)

THE EFFICACY OF A SHORT FORM OF THE MMPI TO EVALUATE DEPRESSION AND MORALE LOSS

ARTHUR CANTER

Johns Hopkins Hospital and School of Medicine, Baltimore, Maryland

Several studies have consistently reported high frequencies in the MMPI of *D* and *Pt* (the so-called 27 or 72 codes) scores among patients suffering severe emotional disturbances, in particular depression (Welsh & Dahlstrom, 1956). One study (Simon & Hales, 1949) found a high incidence of this pattern in hospitalized suicidal veterans. A different study demonstrated a significant relationship between *D* and *Pt* and low self-acceptance (Zuckerman & Monashkin, 1957). It would seem that these two scales alone would be of specific value in outpatient clinics or other situations where brief patient contacts might preclude the use of the full MMPI. Arguments against using single scales of the MMPI as tests in themselves rest upon several grounds. For one thing, some studies indicate that single scales of the MMPI tend to show little validity when tested against standard diagnostic categories (Welsh & Dahlstrom, 1956). Also, there is the possibility that there may be different response tendencies if the scale is isolated from its context of being imbedded among a large host of other scales. While these arguments are compelling, it is conceivable that the discriminability of the various MMPI scales differs from one scale to another and would vary according to the kind of behavior one would try to predict. Whether an isolated scale or two used as a complete test produces less discrimination than the same scale included among a group of others is an empirical matter yet to be convincingly demonstrated. The evidence on hand is consistent in that elevations in both *D* and *Pt* are frequently found in the MMPI profiles of patients having symptoms of depression and loss of self-confidence independent of specific

diagnoses. Thus if one were interested in the identification of individuals who are seriously depressed without regard to their *diagnostic* status, these two keys used alone in a single scale might be expected to work well.

The purpose of this study was to test the discrimination power of a brief form of the MMPI on population samples differing in degrees of depression using only those items necessary to score *D* and *Pt*.

The experimental scale, to be referred to as ML (for Morale Loss), consisted of the 116 items of the MMPI that are scored for *D*, *Pt*, and *K*. The *K* items were included since they enter into a corrected *T* score for *Pt* and also its role as a "suppressor variable" has possible clinical as well as research value. The ML scale was made up in a booklet form allowing it to be used either with an answer sheet or by having the subject (*S*) record his answers directly in the booklet. The average time for individual administration was between 15 and 20 minutes.

Three groups of *Ss* were tested by the ML scale. Group A consisted of 18 patients (nine female) who had been hospitalized because of a neuropsychiatric disorder and who had recently (within a period of 15 days prior to testing) made an unsuccessful active suicide attempt. In every instance some action was taken which was judged to represent a bona fide attempt at self-injury or destruction and not mere verbal threats.¹ The Group A patients were all given a psychiatric diagnosis during the hospitalization following the suicide attempt: five patients were considered to

¹ These included: gunshot wounds, slashing of face and neck, wrist slashing, ingestion of known poisons, overdose of sedatives, inhalation of gas, and attempt at hanging self.

Table 1

Mean Scores^a on the ML Form of the MMPI of Patient and Normal Adult Groups

Group		<i>D</i>	<i>Pt</i>	<i>Pt + K</i>	<i>K</i>
A patients (<i>N</i> = 18)	Mean	80.94	72.17	79.89	11.11
	<i>SD</i>	14.6	15.4	15.9	3.4
B patients (<i>N</i> = 18)	Mean	68.33	56.56	65.67	15.94
	<i>SD</i>	14.9	10.8	11.7	5.2
C normals (<i>N</i> = 18)	Mean	48.72	40.83	47.44	17.94
	<i>SD</i>	7.9	4.4	5.2	6.9

^a Entries for *K* are raw scores. All others are *T* scores.

be schizophrenic; three, manic-depressive depressed; and the remainder severe psychoneurotic reactions with emphasis placed on the depressive symptoms.

Group B contained 18 patients (nine female) who were either hospitalized for a psychiatric disorder or seen in the outpatient department of Henry Phipps Psychiatric Clinic for diagnostic study and treatment. None had a history of overt suicide attempts nor were they considered suicidal at the time of study. The diagnoses included: four schizophrenic reactions, four depressive reactions (not psychotic), and the remainder psychoneurotic reactions with anxiety and some depressive features. Thus the A and B patients differed essentially in the fact that the former had attempted suicide and were considered more severely depressed.

Group C consisted of 18 presumably normal adults (nine female) who had no history of mental illness requiring psychiatric treatment and were comparable to the A and B patients with respect to age and socioeconomic status. The C normals were obtained from groups of persons working in diverse occupations in and around a hospital setting, such as machinists, plumbers, clerks, laboratory technicians, etc. The mean age for Group A was 32.2; for Group B, 31.4; for Group C, 33.3. All ML scales were administered in individual sessions with the Ss. The directions for the test were essentially the same as those used in the standard MMPI booklet.

The *D*, *Pt*, and *K* raw scores were determined from the appropriate MMPI keys and the *T* score equivalents for the two clinical scales were determined from the MMPI

tables. Both the *K*-corrected and uncorrected *T* scores for *Pt* were recorded. The *T* scores of the clinical scales were evaluated in preference to the raw scores since half the Ss were female, and the *T* score norms for *D* and *Pt* have different values for the sexes.

Results and Discussion

The mean scores for the three groups on the *D*, *Pt*, and *K* scales may be seen in Table 1. The significance of differences among the mean scores of the two patient groups and the normal adults was evaluated by simple analysis of variance. Differences between adjacent group means were then analyzed by the *t* test, using the within groups variances obtained from the analysis of variance to determine the best estimate of the standard error of the difference. Table 2 provides a summary of the results of the analysis. It may be seen that both patient groups were significantly higher on *D* and either *Pt* variable, with Group A scoring significantly higher than Group B, considering the patient groups. Only in the case of *K* was there no significant difference found between Groups B and C.

It is apparent that the short MMPI scale as used in this study does permit discrimination between the two types of patient groups which differed originally in the degree of depression, as well as between the patient and normal adult groups. The results also tend to support the findings of Simon and Hales (1949) on their suicidal group in that a higher proportion of the A patients scored

Table 2

Summary of Tests of Differences Among Patient and Normal Groups on *D*, *Pt*, and *Pt + K* *T* Scores

Variable	Values of <i>F</i> for Differences among all Three Groups	Values of <i>t</i>	
		Group A vs. B	Group B vs. C
<i>D</i>	26.88**	2.85*	4.43*
<i>Pt</i>	33.55**	4.09**	4.12**
<i>Pt + K</i>	39.59**	3.89**	4.98**
<i>K</i> (raw scores)	11.78*	3.35*	1.39

* *p* < .01.** *p* < .001.

Table 3

Distribution of the ML Key Scores of Patient and Normal Groups

Range of Scores	Group A (N = 18)	Group B (N = 18)	Group C (N = 18)
29-30			
27-28	////		
25-26	///		
23-24	//		
21-22	////	/	
19-20	//		
17-18		//	
15-16	/		
13-14	//	////	
11-12			
9-10		///	/
7-8		/	/
5-6		///	/
3-4		////	////
1-2			//////
0			//////

above *T* score 70 on both *D* and *Pt* than did the B patients. Seventy-two per cent of the A patients scored this high on both scales as compared to 22% of the B patients. The findings with respect to *K* were consistent with those of other clinical studies in that the

value of *K* drops as the degree of disturbance increases. The *Pt* scale differentiated the groups as well as *Pt* + *K*, although the latter yielded higher *T* scores for each of the three groups.

The fact that *Pt* shows high positive relationships with *D* consistently and appears to be sensitive to emotional disturbances even where no dominant obsessive-compulsive or phobic syndromes are apparent, raises the point that *Pt* measures emotional disturbance per se better than "psychasthenia." Support for this view was provided when the data of the present study were analyzed for the most differentiating of the ML test items that would help separate Group A patients from B patients and both from normals. A total of 30 items seemed to approximate an arbitrary criterion whereby a response in one direction to an item would be given by two thirds of Group A to one third of Group B and less than this by Group C.² Table 3 summarizes

² These items with their MMPI booklet numbers and the direction scored for ML are as follows: 5T, 8F, 32T, 39T, 41T, 67T, 76T, 86T, 84T, 106T, 107F, 145T, 153F, 182T, 189T, 236T, 238T, 290T, 301T, 337T, 340T, 343T, 352T, 357T, 358T, 359T, 361T, 362T, 366T, 383T.

Table 4

Distributions and Means of ML Key Scores of Brucellosis and CMI Groups

Range of ML Scores	I Chronic Brucellosis Group (N = 9)	II Nonchronic Brucellosis Group (N = 8)	III High-Positive CMI Group (N = 15)	IV Low-Negative CMI Group (N = 26)
20+				
18-19			///	
16-17				
14-15				
12-13	/		//	/
10-11	//		///	
8-9	//			/
6-7	//		////	//
4-5	/	/	//	/
2-3		//		////
0-1	/	//	/	//////
Mean ML Score	7.4	2.7	11.3	3.5
I vs. II $t = 3.41$ ($p < .01$)			III vs. IV $t = 5.85$ ($p < .01$)	

the distributions of the scores obtained by such a key from the three groups. When the items were identified as to which of the three MMPI keys were involved, it was found that 22 of the 30 items are scored for *Pt*, 7 are from *D* (with 2 scored oppositely), and 1 from *K* (also scored oppositely).

The likely meaning applied to *Pt* or the key derived as above was explored further by the results obtained from other studies carried out by the writer. In one study (Imboden, Canter, Cluff, & Trever, 1959) both *D* and *Pt* were found to differentiate between two groups of adults who had suffered acute brucellosis infections. One group had recovered completely from the illness while the other was chronically ill for at least several years and up to the time of the study. While the chronic group scored higher on *D* and *Pt*, the psychiatric studies did not suggest prominence of obsessive-compulsive or phobic symptoms. However, the psychiatric studies and a self-concept scale indicated they were depressed and suffering low self-esteem. The records of both groups were scored by the tentative key, labeled ML, made up of the 30 items referred to above.

In a separate study still in progress, Ss who had taken the Cornell Medical Health Questionnaire (Brodman, Erdman, & Wolff, 1956) were divided into two groups and compared on the ML key. One group had extremely high CMI scores indicative of positive medical illness histories, while the other group represented a sample having extremely low CMI scores. Table 4 gives the results of comparing the various groups in these two studies on the ML key. As can be seen, the ML variable separates the groups as might be expected if one assumes that chronically ill or frequently ill people are more likely to be low in self-esteem or morale than healthier people.

From such findings, the high proportion of *Pt* items in the ML key, and other studies in the literature with respect to *Pt* (Welsh & Dahlstrom, 1956), it would appear that the

Pt scale might be profitably reanalyzed with respect to its self-esteem components and those more uniquely associated with psychasthenia. The ML key is offered as only one tentative approach to this problem.

Summary

A short form of the MMPI made up of the *D*, *Pt*, and *K* items only was administered to two patient groups, which differed in their degree of depression and emotional disturbance as judged by their histories of suicidal attempts, and to a group of normal adults. The scores of the three groups on the three keys revealed that they could be differentiated from each other according to the degree of disturbance. Thus the effectiveness these keys have as ordinarily used in the entire MMPI scale was maintained when isolated in this fashion. The *Pt* scale was given further attention when a key made up of the most discriminating items of the short test was found to be heavily weighted with *Pt*. The results of follow-up studies and the long observed relationship between *Pt* and *D* suggest that *Pt* should be redesignated as a morale loss or self-esteem scale. An alternate key to *Pt* was offered by the study toward this end.

REFERENCES

- BRODMAN, K., ERDMAN, A. J., JR., & WOLFF, H. G. *The Cornell Medical Index Health Questionnaire*. (Rev. ed.) New York: Cornell Univer. Med. Coll., 1956.
- IMBODEN, J. B., CANTER, A., CLUFF, L. E., & TREVER, R. W. Brucellosis, III: Psychologic aspects of delayed convalescence. *AMA Arch. Intern. Med.*, 1959, 103, 406-414.
- SIMON, W., & HALES, W. M. Note on a suicide key in the MMPI. *Amer. J. Psychiat.*, 1949, 106, 222-223.
- WELSH, G. S., & DAHLSTROM, W. G. (Eds.) *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. of Minnesota, 1956.
- ZUCKERMAN, M., & MONASHKIN, I. Self-acceptance and psychopathology. *J. consult. Psychol.*, 1957, 21, 145-148.

(Received January 5, 1958)

THE EXPRESSION OF HOSTILITY AND GUILT IN MELANCHOLIC AND PARANOID WOMEN

T. M. CAINE¹

Runwell Hospital, Wickford, England

That feelings of *hostility* and *guilt*, with the concomitant expression of extrapunitive and intropunitive attitudes, are important factors in diagnosis and psychopathology has been recognised for many years. The Freudian doctrine of superego formation, for example, involves aggressive and guilt components. In discussing symptomatology in general psychiatry, Henderson and Gillespie (1950) consider that certain delusions (the idea that one's thoughts are being read, and self-accusatory ideas of delusional intensity) may rest on an affective basis of guilt. In clinical psychology, Foulds (1958) has adapted a technique, devised by Rosenzweig, Clarke, Garfield, and Lehndorff (1946), which elicits the expression of intropunitive and extrapunitive attitudes for diagnostic purposes. Fundamentally, Foulds' adaptation measures the degree to which one is more critical or less critical of oneself than of others. Siegel (1956) has developed a "hostility" scale from the Minnesota Multiphasic Personality Inventory (MMPI) which he has related to "authoritarianism." He found that groups scoring highly on his dimension of authoritarianism had greater *overt* hostility as measured by his hostility scale. In a further paper (Siegel, Spelka, & Miller, 1957) presented at the American Psychological Association Convention in September 1957, evidence was provided of the feasibility of measuring varying aspects of hostility such as "extrapunitive," "intropunitive," and "projected." Alexander (1949) has distinguished between provocative hostile behaviour, projected hostility, guilt

feelings, and inferiority feelings. It seems clear then, that *hostility* and *guilt* are global terms and that different authors have separated out certain aspects for study.

In line with this tradition, the present study is concerned with the measurement of some aspects of *hostility* and of *self-criticism/guilt* at various levels of expression. Since interest was centered on measurement, persons known to be extreme to a pathological degree in the expression of these feelings, namely melancholics (guilty) and paranoid states (hostile) were used as subjects (Ss).

Method

Measurement of Extrapunitive and Intropunitive Attitudes

The purpose, in testing, has been to work from an obvious, overt statement of attitude, to an indirect, covert expression.

At the superficial level, two hostility scales were made up from the MMPI, one consisting of 21 *acting out* hostile statements such as: "at times I have a strong urge to do something harmful or shocking," "I can easily make other people afraid of me, and sometimes do for the fun of it," and "horses that don't pull should be beaten or kicked." A second scale, implying criticism of others and accusing others of hostility to oneself, was devised as a measure of *projected* hostility. This scale consisted of 26 items such as, "I think most people would lie to get ahead," and "I believe I am being plotted against."

As a measure of *self-criticism/guilt* at this level, items implying criticism of self or of admission of guilt were drawn from the MMPI. This scale included such statements as: "I have one or more habits which are so strong that it is no use fighting against them," "I

¹The author would like to express his indebtedness to G. A. Foulds for his encouragement and advice during the course of this investigation, and his thanks to R. Ström-Olsen for permission to publish this report.

have several times given up doing things because I thought too little of my ability," and "I believe my sins are unpardonable." The items of this scale were allocated by the author and agreed upon by the senior psychologist at this department.

At a less direct level, a development of a method of measurement described by Watson, Pritzker, and Madison (1955) was employed. This method consisted of 60 sets of scrambled words which could be assembled in either of two ways by leaving out one word in each set. One completion led to a hostile sentence and the other to a neutral. These authors found that a neurotic group made significantly more hostile completions than did a nonneurotic control group. In the test constructed for the present study, 48 sets of four scrambled words were compiled. Paralleling the MMPI scales, in 16 of the sets one completion led to an acting out hostile response and the other to a neutral; in 16, one completion led to a projected hostile response (as defined above for the MMPI scale) and the other to a neutral, and in 16, one completion led to a self-critical/guilt sentence and the other to a neutral one. Examples:

acting out hostility

HIT YOU
I'LL TAKE

projected hostility

THEY'D ME
CHEAT FIND

self-critical/guilt

COURAGE NO
BOOKS I'VE

The order of the words was randomized in making up the sets. The number of possible combinations was so great that the effect of a methodical approach by the S, such as always reading along the top line from left to right, was reduced to a minimum.

In administering this Sentence Building test, the card with the four scrambled words on it was laid in front of the S, who was instructed to give the first sentence he could make up as fast as he could, by leaving one word out. The emphasis was on reaction time, with a stop watch prominently displayed and each time noted down. The number of acting out hostility constructions, projected hostility,

and guilt construction given first was noted for each S. This may be regarded as a projective type of test in which perceptual factors such as the "selection of personally relevant cues," to use Bruner's (1951) phrase, may be operative. The word or phrase frequency effect, the problem first raised by Solomon and Howes (1951), should be the same for both groups.

At a more indirect level still, eight Thematic Apperception Test cards (TAT), namely 4, 3, 13MF, and 18GF of the standard set, and four specially prepared pictures were used. These cards were selected because they were thought to suggest aggressive themes. Administration was that described by Foulds (1953), involving the four questions, "what is the situation," "what led up to it," "how are they thinking and feeling," and "how does it finish up?" The protocols were then scored for hostility and guilt content by two psychologists independently and without knowledge of the Ss' identity.

A detailed scoring system was devised for this purpose. For the hostility scoring, the mention of murder or death through accident or war was given a score of 3. Stories of physical assault, injuries dangerous to life, or rape were given scores of 2. Minor injuries, verbal assault, or the mention of interpersonal frictions were given scores of 1. For the guilt scoring, the mention of suicide scored 3, remorse with reparation 2, and remorse without reparation 1. If several of these themes were mentioned in one picture the highest score was taken. These highest scores were then summed for the series of eight pictures to get a total score. Since a correlation of .98 for ratings of hostility and .95 for guilt were obtained, the ratings of the senior psychologist were accepted as working scores.

These unusually high correlations may be explained by the avoidance of interpretation. The raters had, in the main, merely to check whether certain statements had or had not been made. Some difficulty arose, however, in deciding whether a particular word or expression could be treated as synonymous with remorse.

At the most indirect and ambiguous level, a tapping psychomotor test was administered. In a study by Foulds, McClelland, McClelland,

and Creasy,² the Tapping test was found to differentiate significantly between paranoid and nonparanoid schizophrenic groups. In a recent investigation by Foulds and Caine (1958), female dysthymic patients were successfully distinguished from hysterics on this

² G. A. Foulds, W. J. McClelland, Marilyn McClelland, and Monica Creasy. Personal communication. 1958.

test. Since, in these studies, hysterics and paranoids scored relatively high on tapping and on various measures of hostility, the Tapping test was introduced here to investigate further its possible relationship to hostility. In the administration here employed the S was merely instructed to tap on a standard sized piece of plain white paper, with a pencil, as fast as he could for 10 seconds.

Table 1

Mean Score Comparisons of Melancholics and Paranoid States on Age, Vocabulary, MMPI Scales, Sentence Building Test, and TAT Hostility and Guilt Ratings

Item	N	Mean	SD	t	p
Age					
Melancholics	17	51.12	6.48	.119	>.1
Paranoids	14	50.78	8.92		
Mill Hill Vocabulary					
Melancholics	17	26.53	5.81	.333	>.1
Paranoids	14	25.86	4.87		
Sentence Building test					
Acting out hostility					
Melancholics	17	7.12	1.62	.088	>.1
Paranoids	14	7.00	4.03		
Projected hostility					
Melancholics	17	5.35	3.25	2.323	<.05
Paranoids	14	8.21	3.36		
Guilt					
Melancholics	17	10.00	2.00	3.733	<.001
Paranoids	14	6.79	2.65		
MMPI hostility					
Acting out hostility					
Melancholics	17	4.88	2.47	.891	>.1
Paranoids	14	5.86	3.44		
Projected hostility					
Melancholics	17	4.47	3.53	3.182	<.01
Paranoids	14	10.07	5.83		
MMPI self-criticism and guilt					
Melancholics	17	15.41	4.06	4.807	<.001
Paranoids	14	7.43	4.87		
TAT					
Hostility					
Melancholics	12	3.50	2.90	2.150	<.05
Paranoids	14	6.36	3.52		
Guilt					
Melancholics	12	1.75	1.64	.079	>.1
Paranoids	14	1.57	1.88		
MMPI depression					
Melancholics	17	84.65	12.74	8.392	<.001
Paranoids	14	53.43	7.18		

This was done three times. The spread of the resulting pencil dots was then determined by counting the number of half inch template squares entered for each trial, and the average for the three taken.

Since the present battery of tests was mainly of a verbal nature a minimum of Grade IV (i.e., above the 10th centile) on the Mill Hill Vocabulary Scale (Raven, 1954) was set. As a diagnostic check the depression scale of the MMPI was included.

Subjects

Three chronic hospitalized paranoids and all paranoid and melancholic female patients fulfilling the verbal level criterion who entered Runwell Mental Hospital during the period September 1957 to February 1958 were considered for testing. Of those entering the hospital, three paranoids could not be included since one refused to cooperate, one was deaf, and one had been leucotomized. Two of the melancholics had to be rejected as untestable. Care was taken not to include any "mixed" cases of paranoids with depressive features or melancholics with paranoid features, since these are to form the basis of a separate investigation. In all, 14 paranoids and 17 melancholics were tested, but only 12 of the latter took the TAT.

Results

It was considered that the distributions of scores on age and tests, with the exception of tapping, were sufficiently normal to justify the use of a *t* test comparison and the results are set out in Table 1. For the Tapping test the nonparametric Mann-Whitney *U* Test (for N_2 between 9 and 20) was used.

It will be seen from Table 1 that there were no significant differences between the groups in age or vocabulary level.

Of the specially devised MMPI scales, the acting out hostility scale failed to differentiate between the groups. On the other hand, the paranoid group had significantly higher scores on the projection of hostility scale and significantly lower scores on the self-criticism/guilt scale. Supporting this finding, at the more covert level of response of the Sentence Building test, the acting out hostility sentences failed to differentiate, but both the

projected hostility and guilt sets have done so in the same direction as the MMPI scales. At the most covert level, the paranoid TATs were given higher hostility ratings than were the melancholics. At this level the guilt measure failed to differentiate. The Tapping test was, surprisingly, the most successful diagnostic test of the battery. The median score of the melancholics on this test was 2, that of the paranoids was 21. The score rankings gave a *U* value of 22, yielding a *p* of $<.002$.

Discussion

This study supports the contentions of Siegel that the varying directions of hostility (if guilt be regarded as introjected hostility) can be objectively measured. Siegel (1957) suggests that individuals may vary more in the manner and direction in which they express hostility than in the amount of hostility they possess, and it may well be that the approach used in the present study could be used to clarify this point. The value of distinguishing various aspects of hostility and guilt has been demonstrated and further logical divisions are indicated if the study were to be extended to other groups. Two further divisions would be to separate a general "criticism of others" from complaints of the direct receipt of hostile acts on the part of others, and a separation of self-criticism (Alexander's inferiority feelings?) from admission of guilt and wickedness.³

The use of different levels of expression might well be a measure of the amount of hostility possessed. Normal, neurotic, and psychopathic groups would be expected to differ from psychotics and between themselves, not only in the aspects of hostility and guilt expressed, but also in the level of expression. One suspects that the consistency at all levels of response might not hold for neurotics and normals, and that this may well indicate the profundity and pervasiveness of the feelings for the individual. Before this can be attempted, however, improvements in the present battery of tests would be required, particularly at the covert levels. It was not possible in the TAT thus administered, for

³ This has now been done for a number of neurotic and psychotic groups. A paper is being prepared for publication.



example, to isolate different aspects of hostility satisfactorily, and since an expression of guilt tended to accompany acts of hostility in the stories, this measure of guilt failed to differentiate between the groups. This may well be more culturally than emotionally determined. It is of considerable interest that although the paranoids avoided acting out hostility on the MMPI scales and Sentence Building test to the same degree as the melancholics, they gave free rein to their feelings on the TAT, in which stories of murder, rape, violence, and quarrelling were marked.

With the Sentence Building test there was considerable loss of distance. Patients often related their response directly to themselves, denying or confirming the truth of a statement made out of the words. That many of the patients were aware of the possibility of alternative responses was obvious from the remarks to this effect, from such comments as, "Oh, I couldn't say that," or exclamations followed by a neutral sentence. It can be argued that the paranoid patient believes that his only way of obtaining release from the hospital lies in convincing the staff of the justification of his position and attitude. An admission of inadequacy, failure, or guilt on his part may seem damaging to his case. Hence his reluctance to voice sentences expressing these feelings, which seem to him, through loss of distance, to have personal relevance. The melancholics, on the other hand, are anxious to impress one with their wickedness, and, indeed, are often worried that the staff may disbelieve their protestations of unworthiness, failure, and guilt. Hence their choice of guilt sentences. A tachistoscopic presentation, to facilitate a closer examination of perception and report with regard to this test, is indicated.

Little can be said of the Tapping test beyond the observation that the group expressing least hostility had lower tapping scores as measured by the number of template squares entered.

An explanation of this finding might be offered, perhaps, in terms of the psychomotor slowing of depressed patients, rather than of hostility. An elaboration of the scoring of this test might be helpful in clarifying this point, since the clinical impression in administering

the test was that force rather than speed had more bearing on the results. Some means of recording the number of taps and the pressure of tapping would be a useful addition.

Summary

Seventeen melancholic patients and 14 paranoid states were given a battery of psychological tests purporting to measure the expression of some aspects of hostility and guilt at various levels of response, from an overt, direct expression of attitude to a covert, indirect one. Consistent differences were found between the groups at all levels. A number of problems worthy of further investigation have been discussed.

REFERENCES

- ALEXANDER, F. *Fundamentals of psychoanalysis*. London: Allen & Unwin, 1949.
- BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & M. N. Ramsay (Eds.), *Perception: An approach to personality*. New York: Ronald, 1951. Pp. 121-147.
- FOULDS, G. A. A method of scoring the TAT applied to psychoneurotics. *J. ment. Sci.*, 1953, 415, 235-246.
- FOULDS, G. A. Superiority-Inferiority Index in relation to frustrating situations. *J. clin. Psychol.*, 1958, 14, 163-166.
- FOULDS, G. A., & CAINE, T. M. Psychoneurotic symptom clusters, trait clusters, and psychological tests. *J. ment. Sci.*, 1958, 436, 722-731.
- HENDERSON, D., & GILLESPIE, R. D. *A text book of psychiatry*. London: Oxford Univer. Press, 1950. Pp. 120, 144.
- RAVEN, J. C. *Guide to using the Mill Hill Vocabulary with Progressive Matrices (1938)*. London: H. K. Lewis, 1954.
- ROSENZWEIG, S., CLARKE, H. J., GARFIELD, M. S., & LEHNDORFF, A. Scoring samples for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1946, 21, 45-72.
- SIEGEL, S. M. The relationship of hostility to authoritarianism. *J. abnorm. soc. Psychol.*, 1956, 52, 368-372.
- SIEGEL, S. M., SPELKA, B., & MILLER, L. The direction of manifold hostility: Its measurement and meaning. *Amer. Psychologist*, 1957, 12, 420. (Abstract)
- SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values and visual duration thresholds. *Psychol. Rev.*, 1951, 58, 256-271.
- WATSON, R. E., PRITZKER, L., & MADISON, P. Hostility in neurotics and normals. *J. abnorm. soc. Psychol.*, 1955, 50, 36-40.

(Received January 6, 1959)

THEORY AND MEASUREMENT OF SOCIALIZATION¹

HARRISON G. GOUGH

University of California, Berkeley

A sociological conception of a continuum of socialization, running from persons of exemplary probity and rectitude at one end, through persons of more typical and less beneficent coadunations of positive and negative propensities, to persons of frankly errant and wayward impulse can easily be delineated. The precise location of a particular individual or group along this continuum may pose something of a problem, but the existential nature of the continuum itself and the fact that persons can and do make ready and reliable judgments in reference to it cannot be gainsaid.

An adequate theory of socialization must pay attention to this continuum and be in consonance with it. This admonition also applies to any procedures of measurement employed in behalf of the theory. That is, a scale of measurement for "socialization" should position individuals in the "asocial," "normal," and "supernormal" zones of the continuum in general accordance with the verdict which the sociocultural environment has handed down concerning them.

If the scale of measurement is psychological, appealing to factors of individual response and calibrated on such reactions, then the problem is a psychosociological one, as it were, requiring the scaling of the psychological dimension in such a way as to covary with the sociological one. Discrepancies are of course to be expected in individual instances between the sociological baseline and the psychological measurement, if for no other reason than that the culture will occasionally

make mistakes in putting some men in prisons and others in positions of trust and responsibility. One of the benefits to be derived from a psychological method for assessing the socialization continuum is that errors of this kind can be identified.

For a number of years the author has been concerned with a theoretical formulation of the socialization problem which attempts to do justice to this notion of the continuous nature of the function (Gough: 1948, 1954; Gough & Peterson, 1952). Perhaps a brief summary of the theoretical position would be in order here. Socialized behavior is behavior based on a proper viewing of the self as a social object, where the terms self and social object are used in the sense defined by Mead (1934). That part of the personality which links an individual to the social community is the "self." The sense of self, or view of self, is a product of social interaction and of the capacity of the individual to view himself as an object; that is, from the standpoint of the other. The self, therefore, has its roots in role taking, in a developing objectification of the critiques and evaluations of oneself. In early life these views of self (called "me's" by Mead) may be discrete and unintegrated, but in time a certain communality and consistency in the patterns permits the evolution of a conception of the "generalized other" which represents social reality as witnessed by the self.

Such a process of introjection of societal standards is always coupled with a degree of uncertainty or independence in the expression of the self. This unpredictable element in the self is called the "I" by Mead. On the other hand, social adaptability, cooperation, and socialization itself are functions of the "me's," that is, of the role taking experiences and the

¹ Earlier work on this program of research was conducted under a series of grants from the National Institute of Mental Health, U. S. Public Health Service. Since 1953 the program has been supported by two basic research grants awarded by the Ford Foundation.

role taking capacities. Thus the degree to which the person will be able to govern internally his thought and behavior in accordance with the imperatives of his culture will be a consequence of the depth and validity of the role taking experiences that he has enacted.

In an earlier paper (Gough, 1948) the author tried to show how this theoretical position could be brought to bear on the problem of psychopathy, and how the various symptomatic expressions of psychopathy could be deduced from the fundamental propositions of role theory. Later (Gough & Peterson, 1952) an attempt was made to develop a measuring instrument for asocial behavior using questionnaire materials derived on the basis of the role theory and focusing on the domain of social interaction and role taking experiences. Diagnostic items pertaining to the self view and role psychology were assembled and then administered to experimental samples, male and female. These included high school students, nondelinquent disciplinary problems in high school, and institutionalized delinquents. Statistical analysis permitted the identification of 64 items (from some 200 evaluated) having significant correlations with the socialization criterion.

These 64 items were grouped into a scale which was then cross-validated on a military sample of 1,092 inductees vs. 99 stockade prisoners, and a second military sample in which 144 prisoners with two or more offenses were compared with 209 first offenders. The difference between the mean of the inductees and that of the prisoners was 7.26, critical ratio 11.52. For the first offenders vs. recidivists the difference between means was 2.45, critical ratio 3.31. The signs in each instance were in the predicted direction, but even with use of the two-tailed test both critical ratios give probabilities well under .01.

However, these two cross-validated citations cover only two points on the socialization continuum: delinquents vs. nondelinquents and first offenders vs. repeaters. It must also be shown that *socialized* behavior can be brought under the purview of the theory and its scale of measurement to the same extent as *asocial* behavior. A scale or theory which would apply only to the distinc-

tion between delinquents and nondelinquents would be a limited one indeed. The goal of the theory, on the contrary, is to encompass the full range of phenomena implied by the continuum and the goal of the scale is to locate persons and groups in their proper places in any of its zones.

To test the validity of the scale in making such differentiations, a greater variety of samples was needed than reported in the 1952 paper, especially samples in the "more socialized" regions of the continuum. Such samples have gradually been obtained over the past seven years, in a sufficient number it would now seem to warrant analysis and reporting.

Shortening of the Scale

As a first step, the 64-item scale developed in the 1952 study was item analyzed against the first offender vs. repeater dichotomy. Such an analysis follows logically from the theoretical conception of socialization-asocialization as a continuum, and from the contention that individual items as well as the full scale should possess validity at different regions of this continuum. Ten items were eliminated by the analysis, leaving 54 in the shortened scale.² The following are representative of those retained (shown with the scored response for socialization).

1. Before I do something I try to consider how my friends will react to it. (true)
2. I often think about how I look and what impression I am making upon others. (true)
3. I would rather go without something than ask for a favor. (false)
4. I find it easy to "drop" or "break with" a friend. (false)

Because the purpose of the scale is to position either individuals or groups along the basic underlying socialization continuum, it has been designated "So" for socialization. A definition of the psychological implication of the scale can be given as follows: "To indicate the degree of social maturity, integrity, and rectitude which the individual has attained."

² This 54-item version is the one included in the California Psychological Inventory (Gough, 1957). The CPI Manual gives a great deal of clinical and psychometric information about the scale.

Table 1

Means and Standard Deviations on the California Psychological Inventory So (Socialization) Scale for Male Samples Indicated

More Socialized				Less Socialized			
Sample	N	M	SD	Sample	N	M	SD
1. Nominated high school "best citizens"	90	39.44	4.95	1. High school disciplinary problems	91	31.25	5.40
2. Medical school applicants	70	39.27	4.82	2. County jail inmates	177	29.27	6.44
3. Banking executives	121	39.00	4.59	3. Prison inmates, New York	94	28.28	5.80
4. Regional wholesale flour salesmen	85	38.31	4.37	4. Young delinquents, California	426	28.07	5.72
5. City school officials	200	37.58	4.19	5. Prison inmates, California	177	27.76	6.03
6. Psychiatric aides	132	37.57	4.37	6. Training school inmates, New York	100	26.53	4.89
7. Business executives	116	37.47	4.72	7. Inmates, federal reformatory	230	26.23	6.53
8. College students	1745	37.41	5.28				
9. Civil Service supervisory personnel	122	37.25	4.98				
10. Electronic technicians	55	36.93	5.66				
11. Correctional officers	620	36.72	5.47				
12. Semiskilled workers	108	36.62	5.17				
13. High school students	4474	36.46	5.95				
14. Social work graduate students	182	36.40	4.62				
15. Military officers	495	36.38	4.74				
16. Machine operators	105	35.99	4.98				
17. Psychology graduate students	142	34.60	4.13				
18. Selective service inductees	139	32.83	6.71				
Total	9001	36.74	5.61		1295	27.98	6.08

Note.— $M_1 - M_2 = 8.76$; $CR = 48.94$; $P < .001$; $r_{bis} = +.73$.

New Samples

The present findings are based upon an entirely new series of 41 research samples, totalling 1,295 male delinquents and 9,001 non-delinquents, and for females 784 delinquents and 9,776 controls. The samples cover a wide spectrum of the socialization continuum, ranging from nominated "best citizens" through various occupational and professional groups, through disciplinary samples, to known delinquents and prison inmates.

Table 1 lists the male samples tested and presents summary statistics on the 54-item So scale. The samples are ranked by mean score on So, with "more socialized" or nondelinquent groups on the left, "less socialized" or

delinquent on the right. The highest mean score among the 25 samples is observed in that of the nominated best citizens, and the lowest in the sample of federal reformatory inmates. In general, the rank ordering of samples by the So scale seems to accord quite well with what would occur if they were ranked sociologically for socialization. It should be noted that all seven of the samples carrying some explicit designation of asocial behavior score below the mean of the lowest scoring of those samples not so designated. Comparison of the 9,001 cases in the more socialized group of samples with the 1,295 cases in the less socialized group gives a difference between the means of 8.76, criti-

cal ratio 48.94, $P < .001$. The biserial correlation with the So scale for this same dichotomy is $+ .73$.

Quite clearly, the So scale locates male samples along the socialization continuum in the way required by the underlying theory. The question now is, can this same differentiation be demonstrated for women? Table 2 presents summary statistics on the CPI So scale for 15 female samples. The progression of mean scores for the female samples in Table 2 shows the same correspondence to the underlying socialization continuum as was previously observed for the male samples. There is again a perfect separation between the nine samples in the more socialized column and the seven less socialized samples manifestly designated, in one way or another, for some defection from the socialization norms. For both the male and female samples the validity of the specific placement of a sample on the socialization continuum by the average So score might be disputed, but the general correspondence between the psychometric ordering and the sociological hierarchy is remarkable. The statistical tests are also significant for the female samples, with a difference between the means of the more

vs. less socialized cases of 9.52, critical ratio 37.83 ($P < .001$), and a biserial r of $+ .78$.

If one is willing to accept these results as indicating the validity of the So scale as a measure of socialization, some interesting observations can be made about the relative standings of the groups tested. For example, there is the bemusing finding that psychologists (University of California, Berkeley, psychologists, it might be cautioned) tend to rank rather low on the socialization continuum, although fortunately not quite so low as to fall into the outrightly troublesome region of the dimension. The sample of unmarried mothers stands midway in socialization between high school disciplinary problems and county jail inmates. In both male and female rankings incarcerated persons score at the low end of the scale. It should also be noted that female samples in nearly all instances score significantly higher than their male counterparts.

The grouping of cases in Tables 1 and 2 into more socialized and less socialized is maintained in Table 3, where percentage frequency distributions are presented. Some indication of possible cutting scores and their screening efficiencies for the usual delinquent

Table 2
Means and Standard Deviations on the California Psychological Inventory So
(Socialization) Scale for the Female Samples Indicated

More Socialized				Less Socialized			
Sample	N	M	SD	Sample	N	M	SD
1. Nominated high school "best citizens"	90	41.51	4.55	1. High school disciplinary problems	87	34.79	7.00
2. High school students	5295	39.69	5.55	2. Unmarried mothers	213	32.92	6.24
3. College students	3452	39.37	5.05	3. County jail inmates	51	29.61	5.86
4. Factory workers	291	38.99	4.76	4. Prison inmates, Indiana	127	28.37	6.24
5. Psychiatric aides	67	38.70	3.91	5. Prison inmates, California	135	28.36	5.68
6. Nurses	142	38.24	4.89	6. Prison inmates, Wisconsin	76	26.83	7.04
7. Airline hostesses	60	38.07	4.51	7. Young delinquents, California	95	25.83	5.13
8. Social work graduate students	320	37.99	4.38				
9. Psychology graduate students	59	36.44	3.93				
Total	9776	39.46	5.30		784	29.94	6.89

Note.— $M_1 - M_2 = 9.52$; $CR = 37.83$; $P < .001$; $r_{bis} = +.78$.

Table 3

Proportion of Cases Scoring at or above the Indicated Scores of the California Psychological Inventory So (Socialization) Scale for Ss Classified as More Socialized and Less Socialized

So Scale Score	Males		Females	
	More Socialized N = 9001	Less Socialized N = 1295	More Socialized N = 9776	Less Socialized N = 784
42	21	1	40	4
41	28	2	47	6
40	34	3	55	9
39	41	4	62	11
38	48	6	68	15
<hr/>				
29	91	47	97	58
28	94	53	98	64
27	95	60	98	68
26	97	65	99	72
25	98	70	99	76
24	98	75	99	81
23	99	81	99	85
22	99	86	99+	88

vs. nondelinquent dichotomy can be gained from this table.

However, one might still ask whether any practical validity of the So scale for making the two-category classification in mass testing might not be vitiated by a base rate limitation. That is, if Category A (delinquency) is quite rare, the rules of inverse probability might make it more valid (accurate) in the long run simply to predict that everyone tested will be non-A. As Cureton (1957) has shown, this problem can be handled by proper setting of cutting scores so as to allow for base rate asymmetries. The optimal cutting score for predicting a dichotomous criterion (delinquent vs. nondelinquent here) from a continuous variable (the So scale) is the point of intersection of the smoothed frequency distributions of the two classes on the predictor variable, where the areas of the two distributions are proportional to the base rates.

The question in our case is what base rate frequencies to use? As an example, suppose it is assumed that the true incidence of delinquent personalities is 10% of the population. Using Cureton's method for smoothing the distribution curves for the four samples in Table 3 and checking for the defined in-

tersections leads to the specification of a cutting score of 23 (and below) for delinquency in males, and of 26 (and below) in females. If the true incidence of individuals sufficiently asocial to be classed as delinquent personalities is at least 1 in 10, then diagnostic classifications of all men with scores of 23 and below as delinquent types, and as nondelinquent types those with scores of 24 and above, will be more accurate than the 90% accuracy level which would automatically follow from classifying everyone tested as a nondelinquent.

With a more misanthropic view that delinquent personalities in reality constitute as much as 20% of the population, the optimum So cutting score for males is at 25 or below for delinquency. For females, the one in five assumption determines a cutting score of 28 (and below).

Perhaps two comments should be interjected before closing this section. The first is that the possibility of deriving dichotomies of the above type does not in any way alter the fundamental theory of the scale and of its psychosociology as representing a continuous socialization function. The second is that specifications for optimum use of the scale in other categorization problems could just as easily be made, for example, for super-citizens vs. all others.

Work of Others

A number of other investigators have carried out studies using the CPI So scale which help to illuminate the relationship between the measure and the theoretical continuum of socialization hypothesized to underlie it. One of these studies is that by Clark Vincent of unwed mothers.³ The So scale data were gathered for 232 subjects (Ss), tested in several public and private maternal care centers. For this sample the mean So score was 32.25, *SD* 6.58. These figures closely approximate those for the smaller sample of 213 reported in Table 2 earlier.

The theoretical rationale for the So scale, however, requires a more detailed analysis of this sample, an analysis attentive to the gradations of the continuum. In the present

³ Permission to use these data was very kindly given by Vincent, Univer. of California, in his book, *Unwed Mothers*, to be published.

instance socialization gradations can be reflected by the number of pregnancies of each S. Classification on this basis gave the following So results:

Group	N	M	SD
1. One child	201	33.60	5.65
2. Two children	18	24.39	5.98
3. Three or more children	13	22.23	4.16

The progression of means here is that specified by the theory of the scale, and is supported statistically by the significance of the *F* test over the three means. Groups 2 and 3 can be combined into a sample of 31 Ss having had two or more illegitimate children. The biserial correlation for this sample vs. the 201 females having only one child is $+ .83$. This coefficient offers rather striking evidence of the power of the So scale to discriminate within a certain range of the continuum.

The mean scores listed above can also be reviewed for hypotheses about the proper location of each sample on the socialization continuum. The Ss having one child score at about the same point as high school disciplinary problems; Ss having two or more children fall into a distinctly lower region of the dimension, into the range populated by the delinquent and criminal samples.

The next study to be considered is that of Donald (1955). He administered the So scale to 230 consecutively admitted federal reformatory inmates. The mean of 27.77 (*SD* 6.53) places the total sample in its proper position on the continuum. The question for analysis is again that of differentiation within the total sample. Five of Donald's comparisons can be reported. The first is one in which socialization theory would *not* predict a difference, a comparison between white and Negro inmates. This expectation was confirmed, as the two part-samples did not differ significantly. The mean for 56 Negro inmates was 29.39, for 174 white inmates 27.75; however, the *t* test of the difference was not significant.

A second of Donald's comparisons was between 135 Dyer Act commitments, $M = 25.96$, $SD = 6.14$, and the remaining 95 cases, $M = 30.34$, $SD = 6.19$. Dyer Act offenses are in general more serious, and this difference is therefore in the predicted direction; the *t*

ratio of 5.3 is significant well beyond the .01 level.

Comparison 3 pitted 111 boys with zero or one previous commitment vs. 119 boys with two or more. The means and standard deviations were 29.72 and 6.27 vs. 25.95 and 6.23. The difference is in the expected direction, and the *t* ratio of 4.5 is again statistically significant.

The fourth contrast was between boys whose first commitment occurred at age 15 or before vs. those whose first commitment came at age 16 or later. Socialization theory would predict lower So scores for the former group. Findings were in accord with these expectations; for the 96 boys in the first group $M = 24.76$, $SD = 5.89$, and for the 134 in the second $M = 29.93$, $SD = 6.09$. The *t* ratio for this difference was 6.4, $P < .01$.

The last comparison is a novel one, contrasting boys committed for "moonshining" infractions vs. all others. Federal liquor law violations lead to federal institution commitment, but from a socialization standpoint seem more akin to troublesome or moderately wayward behavior than to the asocial nature of most felonies. Therefore, one would expect higher So scores for this special sample. Such indeed was the case, for the 18 "moonshiners" had a mean score of 32.44, SD 6.99, and the 212 remaining inmates had a mean of 27.37, SD 6.33. This difference is also statistically significant, the *t* ratio being 2.9. The absolute level of the So mean, 32.44, should also be noted. It would rank the "moonshiners" just above "high school disciplinary problems" on the socialization hierarchy (see Table 1).

The third study to be mentioned is that of the Pilot Intensive Counseling Organization Project (1956) being conducted at the Deuel Vocational Institution in California under the auspices of the California Youth Authority and the California State Department of Corrections.⁴ In the PICO project, youth authority wards are assigned on a random basis, at intake to experimental and control samples, the experimental cases participating in an intensive counseling program. The aim of the

⁴The data reported here were very generously made available by Sherwood Blair, PICO project supervisor, and Alvin Rudoff, research coordinator.

project is to discover the effects of this counseling program, hoping to find that the experimental cases will show better parole and postinstitutional records. The data reported here are all drawn from cases in the control sample.

The CPI is administered to all Ss in the study at the time of admission to the training school. From the project records three samples were established for our use: (1) parolees receiving suspensions during the first four months of parole; (2) parolees suspended during the fifth or a later month; and (3) parolees still free of suspension after five or more months on parole. The presumed socialization gradient is from lowest (1) to highest (3). CPI So scale statistics for these samples are as follows:

Sample	N	M	SD
1. Suspended within four months	90	27.26	5.10
2. Suspended after four months	90	28.18	5.36
3. Five or more months of parole without suspension	226	28.25	5.78

The progression of mean scores is in the direction specified, but the F test falls short of statistical significance. The follow-up period in the PICO project was not far enough along at the time of this writing to permit designation of a "successful discharge from parole" sample; the third group above is only an approximation of such a sample. If comparison is limited to early suspensions on parole (Sample 1) vs. later suspensions on parole (Sample 2), a critical ratio of 1.18 for the difference between the two means is obtained. Using the one-tailed test of significance, appropriate here because the difference is in the predicted direction, the probability level is between .05 and .10.

The fourth reference to the work of other investigators concerns the program of studies being conducted by Reckless and Dinitz and their associates at the Ohio State University (Dinitz, Kay, & Reckless, 1957; Reckless, Dinitz, & Kay, 1957; Reckless, Dinitz, & Murray, 1957). In the Reckless, Dinitz, and Murray study, 125 boys living in high delinquency areas in Columbus, Ohio, were nominated by teachers as being "insulated" against delinquency. These nominations were confirmed by a brief social service check of

school and police records. The mean So scale score for this sample was 39.43, *SD* 6.42. Only one sample in Table 1, the high school best citizens, had a higher average score. In another study (Reckless, Dinitz, & Kay, 1957), these 125 boys were compared with 101 boys from the same high delinquency areas nominated by their teachers as being delinquency prone. The So scale statistics for this second sample were $M = 31.40$, $SD = 7.99$. The difference between this and the earlier mean was highly significant.

The paper by Reckless, Dinitz, and Murray (1957, p. 568) also contains a comment bearing directly on the view of socialization as a continuous dimension and on the capacity of the So scale to differentiate within regions of this dimension:

The delinquency vulnerability and social responsibility scales did more than differentiate between the potentially delinquent and non-delinquent nominees. The scales also discriminated significantly, within the sample of potential delinquents, between those who had and those who had not experienced previous police and court contact.

The "delinquency vulnerability" scale referred to in the quotation is the CPI So scale discussed throughout the present paper.⁵ The "social responsibility" scale is another one of the scales included in the full set of 18 in the California Psychological Inventory.

Summary

Arguments in behalf of viewing socialization as a continuous dimension rather than as merely a dichotomy of social vs. asocial behavior were presented. A theory of socialization was summarized, and a method of psychological measurement consonant with this theory was described.

The greater portion of the paper was then taken up with findings seeking to demonstrate the validity of the measuring scale along the full range of the socialization continuum.

Evidence from the writer's inquiries as well

⁵ In the Ohio State study the scale was scored in the opposite direction, so as to emphasize the delinquency vulnerability implications. The mean values reported in the present paper, however, have been reconverted to So scale form, so that higher scores are indicative of higher socialization levels.

as from those of other investigators gave strong support to the systematic validity of the scale.

REFERENCES

- CALIFORNIA YOUTH AUTHORITY AND STATE DEPARTMENT OF CORRECTIONS. Technical Report Series. Pilot Intensive Counseling Organization Project. Deuel Vocational Institute, Tracy, California, 1956, 1958.
- CURETON, E. E. Recipe for a cookbook. *Psychol. Bull.*, 1957, 54, 494-497.
- DINITZ, S., KAY, BARBARA, & RECKLESS, W. C. Delinquency proneness and school achievement. *Educ. res. Bull.*, 1957, 36, 131-136.
- DONALD, E. *Personality scale analysis of new admission to a reformatory*. Unpublished master's thesis, Ohio State Univer., 1955.
- GOUGH, H. G. A sociological theory of psychopathy. *Amer. J. Sociol.*, 1948, 53, 359-366.
- GOUGH, H. G. Systematic validation of a test for delinquency. *Amer. Psychologist*, 1954, 9, 381. (Abstract)
- GOUGH, H. G. *Manual for the California psychological inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1957.
- GOUGH, H. G., & PETERSON, D. R. The identification and measurement of predispositional factors in crime and delinquency. *J. consult. Psychol.*, 1957, 16, 207-212.
- MEAD, G. H. *Mind, self, and society*. Chicago: Univ. Chicago Press, 1934.
- RECKLESS, W. C., DINITZ, S., & KAY, BARBARA. The self component in potential delinquency and potential non-delinquency. *Amer. sociol. Rev.*, 1957, 22, 566-570.
- RECKLESS, W. C., DINITZ, S., & MURRAY, ELLEN. The "good" boy in a high delinquency area. *J. crimin. Law Criminol.*, 1957, 48, 18-25.

(Received January 12, 1959)

A COMPARISON OF QUANTIFIABLE RORSCHACH ANXIETY INDICATORS IN HYPNOTICALLY INDUCED ANXIETY AND NORMAL STATES

EUGENE E. LEVITT

Indiana University Medical Center

AND HANUS J. GROSZ

Albert Einstein College of Medicine, New York

Most Rorschachers believe that the Rorschach test is a fairly sensitive indicator of anxiety in the subject (S). However, there is much disagreement on specific anxiety signs, with the possible exception of the shading response. Eichler (1951a) listed 15 factors, which one or another expert alleged to be an indicator, but almost every Rorschach factor which can be quantified has been employed in this fashion in some investigation. The usual design of such a study is to contrast performances of groups of psychiatric patients and normal individuals. In general, these studies have tended to produce equivocal and often conflicting results. Interstudy differences in administration and scoring techniques may have caused these discrepancies in part, but the most probable confounding factor is sampling. There appears to be so many characteristics of the S which can influence Rorschach performance that the matching of patients and normals to eliminate bias is extremely difficult to accomplish. Successful matching is likely to be accomplished by pure accident, and the experimenter cannot be expected to know when this has occurred.

The most obvious way around the sampling problem is to use the same Ss for both the normal and anxiety performances. This introduces a new problem: how to induce anxiety in normal individuals. Many techniques have been used (cf. May, 1950) but their efficacies are usually questionable, if not downright dubious.

The use of hypnotic suggestion to produce

emotional states in normals has been reported periodically in the literature, and recently by Eichhorn and Tracktir (1955a, 1955b). The results suggest that this approach is relatively successful, and that it could be utilized to induce anxiety.

Procedure

The basic design of the study was to compare the Rorschach records of normal Ss as they would be obtained under ordinary circumstances with those obtained in an hypnotically induced anxiety state. In order to control for the effects of an hypnotic state itself on test performance, a record in hypnosis without anxiety was also obtained.

The Ss were 12 medical and nursing students, six males and six females. All Ss were volunteers and were paid for participation in the study. They had been selected from a pool of such Ss on the basis of two criteria: capacity to develop amnesia for performance in hypnosis, and general psychological adjustment. All Ss scored within usually accepted normal limits on the Taylor Anxiety Scale and the Barron Ego Strength Scale and were screened by psychiatric and medical examinations.

The three records were obtained from each S on a single day in the following order: in the hypnotic condition, in hypnotically induced anxiety, and, finally, in the waking state. The S was made amnesic for the blots after each of the first two performances.

Diffuse, nonspecific anxiety was induced in each S by a psychiatrist (H. J. Grosz) using a detailed suggestion in which the words "anxiety," "fear," "apprehension," and "panic" were used. This use of a number of general synonyms for anxiety represents an effort to circumvent the effects of different private meanings of the single word "anxiety" among the Ss. The Ss were told that they did not know what was making them anxious. After the induction of anxiety, all Ss verbalized anxiety feelings and all showed some physical manifestations commonly as-

Table 1
Factors Treated by Covariance Analysis
(Influence of *R* Removed)

Factor	<i>F</i> ratio	<i>r</i> with <i>R</i>	Adjusted Condition Means		
			Waking	Hypnosis	Anxiety
<i>W</i>	6.49**	-.10	7.16	9.81	7.36
<i>Dd</i>	1.02	.56**	2.12	1.85	2.71
<i>F</i> (- <i>FM</i>)	2.88	.71**	15.19	14.17	11.80
<i>F</i> (+ <i>FM</i>)	1.71	.79**	17.89	18.45	16.23
<i>F+</i>	5.42*	.65**	8.63	8.51	6.02
<i>M</i>	0.12	-.04	3.60	3.41	3.64
<i>FM</i>	5.36*	.23	2.70	4.28	4.51
<i>Sum Y</i>	7.82**	.75**	2.52	2.23	4.05
<i>Sum C</i>	0.51	.34	5.24	5.99	5.91
<i>V</i>	0.06	-.25	1.73	1.79	1.96
<i>S_p</i>	0.88	.29	3.23	4.13	3.72
<i>P</i>	1.83	.50*	6.39	6.77	5.52
<i>Tot H</i>	2.05	.52*	5.43	5.48	7.26
<i>Tot A</i>	0.49	.62**	10.04	11.23	10.73
<i>Anat</i>	0.45	.63**	2.15	2.80	2.55
<i>FY</i>	1.83	.73**	3.83	3.75	4.91
<i>FC</i>	1.97	.28	4.17	2.96	3.37
<i>CF</i>	1.65	.24	3.03	4.12	3.43
<i>Hd</i>	1.57	.58**	1.74	2.11	2.90
<i>Ad</i>	0.45	.33	2.24	1.72	1.96

* Significant at the .05 level.

** Significant at the .01 level.

sociated with anxiety, such as trembling, agitation, pallor, etc. Ratings of the degree of anxiety were made independently by the two authors for each *S*, using a five-point scale on which a rating of one represented zero anxiety as in the waking state. The interrater agreement, as reflected in a tau coefficient of .47, is not high, but the raters did agree that every *S* showed some anxiety above what they had previously manifested. The mean rating by the psychiatrist was 3.04, *SD* = .82. The psychologist's mean rating was 2.92 with an *SD* of .83.

Results

A total of 25 Rorschach factors were analyzed. For 20 of these, the treatment was a covariance analysis in which the influence of *R*, the number of responses, was held constant.¹ The *F* ratios among mean scores for

¹ The use of a covariance analysis to control for the effects of *R* has been suggested by Eichler (1951b) as the most effective, economical method when *N* is small. Since the distributions of many Rorschach factors obtained from unselected samples as well as from many types of selected samples often are markedly skewed, a transformation is ordinarily required before applying the covariance approach. This was not true in the present study. Most of the distributions tended to be relatively symmetrical

the three conditions, the correlation with *R*, and the adjusted condition means for these 20 factors appear in Table 1. Three factors—*R*, *F+*%, and median reaction time of first response—were treated by variance analyses; the data are in Table 2. There were several

about a maximum ordinate near the center of the distribution. Means and medians coincided roughly. A transformation was therefore not indicated. The primary explanation of the absence of skewness apparently lies in the relatively limited intelligence range in our sample. Most Rorschach variables, especially determinants, show some relationship to intellectual factors; this is reflected in their correlations with *R*, as shown, for example, by Fiske and Baughman (1953). *R* is itself heavily dependent upon intelligence; as Beck (1954, p. 53) remarks, "the response total is one of the test's best indices of intelligence in liberation." The relatively high mean *R* in our group is a second explanation of the absence of skewness. This mean would fall at about the 70th percentile of the *R* distribution for either the normals or patients in the Fiske and Baughman study. What this means, simply, is that our *Ss* produced more scores above 2 on various factors, and fewer of the scores of 0 and 1 which form an important causal component of the skewed distributions usually obtained with Rorschach factors.

factors whose incidence was too low and whose distributions were too skewed to permit covariance analyses. Two of these, *M*— and pure *C*, were components of broader factors which were found in the covariance analyses to be unrelated to *R*. These factors were analyzed by means of the Cochran *Q* Test (Siegel, 1956) for correlative proportions with more than two conditions. These results are shown in Table 3; the figures under the respective conditions show the number of *Ss* whose records had at least one of the factors.

There are significant differences among the three conditions on 6 of the 25 factors: *W*, *F*+, *FM*, *Sum Y*, median reaction time, and *M*—. The inferences are clear-cut with respect to *F*+, *Sum Y*, reaction time, and *M*—, since in each of these three conditions the significance is a function of a difference between the anxiety state on one hand and the other two states on the other. In the case of *W*, the conclusion is that its incidence is facilitated by hypnosis and is then depressed by the anxiety condition. We may conclude, then, that as far as the results of this study are concerned, these five indices reflect anxiety. In the case of *FM*, the significance is due to a common increase in both hypnosis and anxiety, with no difference between the two. The inference must be that the increased incidence is a result of the hypnotic state itself, and not of anxiety. It is of incidental interest that the *F*+ which is significant is *not F*+ % in the conventional sense but *F*+ as a function of the total number of responses. The conventional *F*+ %, which appears in Table 2, was not significant, though it is in the right direction. Of the 15 factors suggested as anxiety indices by Eichler (1951a),

Table 2

Factors Treated by Variance Analyses

Factor	<i>F</i> ratio	Condition Means		
		Waking	Hypnosis	Anxiety
<i>R</i>	1.61	34.3	31.0	29.8
<i>F</i> + %	1.09	89	87	84
Mdn. <i>RT</i> (secs.)	4.83*	24.4	18.0	31.2

* Significant at .05 level.

Table 3

Factors Treated by Cochran *Q* Test

Factor	Chi Square	Factor Incidence (<i>N</i> of <i>Ss</i>)		
		Waking	Hypnosis	Anxiety
<i>M</i> —	8.40*	1	2	6
Pure <i>C</i>	3.43	1	3	5

* Significant at .05 level.

12 appear in this study. Of these, two—*W* and *Sum Y*—significantly differentiated the anxiety state (interestingly enough, this was also Eichler's result). Four others—*R*, *Dd*, *Hd*, and *P*—are in the right direction. The remainder—*M*, *Sum C*, *Ad*, *Total A*, anatomy responses, and *F*—are unpredictable. Of interest is the fact that *F*, as in the Eichler study, seems to be reduced rather than increased in anxiety. Also as in the Eichler study, this seems to be a function of the increase in shading responses. The data of Table 1 also show that it makes very little difference whether *F* is scored to include animal movement responses à la Beck, or without them as in the Klopfer system. However, the fact that *FM* seems to be affected by hypnosis suggests that it has some merit in the general analysis of personality.

About half of the factors treated by covariance analyses are significantly related to number of responses, despite the small number of *Ss*. This is, of course, not unexpected. Fiske and Baughman (1953) reported that all factors were related to *R* in a large sample of outpatients, and all but a few in the Spiegel normal sample on which Beck's (1950) norms are based. The absence of correlation of many variables in our study is probably due to the attenuated range of *R*, which is itself due to the limited intelligence range in the sample. It is interesting to note that the correlation of .50 between *R* and number of populars is very close to the coefficient of .48 reported by McCall and Doleys (1955) for the Spiegel sample.

Even though our results are based on a small group of *Ss*, they may nevertheless be more revealing than those of larger scale, con-

trasting groups studies. Despite other limitations, there is no sampling bias involved in the selection of the "contrasting groups" in our investigation. Whether the use of individuals of relatively high intelligence and hypnotic susceptibility restricts generalizations from our results is a question requiring further study. Certainly the mean scores are not representative. This does not, however, mean that the *change* in mean scores from state to state might not be characteristic of a broader population.

Summary

Rorschach records were obtained from 12 carefully selected, normal individuals in the waking state, in hypnosis, and in hypnotically induced anxiety. Appropriate analyses of 25 quantifiable Rorschach factors indicate that decreases in *W* and *F*+, and increases in *Sum Y*, reaction time of first response, and incidence of *M*- reflect the anxiety state. An increase in *FM* was found to characterize the hypnotic state per se, but not the anxiety state.

REFERENCES

- BECK, S. J. *Rorschach's test: I. Basic processes* (2nd ed.) New York: Grune & Stratton, 1950.
- BECK, S. J. *Rorschach's test: II. A variety of personality pictures*. New York: Grune & Stratton, 1954.
- EICHHORN, R., & TRACKTIR, J. The effect of hypnotically induced emotions upon gastric secretion. *Gastroenterology*, 1955, 29, 432-438. (a)
- EICHHORN, R., & TRACKTIR, J. The relationship between anxiety, hypnotically induced emotions and gastric secretion. *Gastroenterology*, 1955, 29, 422-431. (b)
- EICHLER, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.*, 1951, 46, 344-355. (a)
- EICHLER, R. M. Some comments on the controlling of differences in responses on the Rorschach test. *Psychol. Bull.*, 1951, 48, 257-259. (b)
- FISKE, D. W., & BAUGHMAN, E. E. Relationships between Rorschach scoring categories and the total number of responses. *J. abnorm. soc. Psychol.*, 1953, 48, 25-32.
- MAY, R. M. *The meaning of anxiety*. New York: Ronald, 1950. Ch. 4.
- MCCALL, R. J., & DOLEYS, E. J. Popular responses on the Rorschach test in relation to the number of responses. *J. clin. Psychol.*, 1955, 11, 300-302.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

(Received January 19, 1959)

THE SPIRAL AFTEREFFECT TEST (SAET) AS A PREDICTOR OF NORMAL AND ABNORMAL ELECTROENCEPHALOGRAPHIC RECORDS IN CHILDREN¹

THEODORE H. BLAU² AND ROBERT E. SCHAFFER

Byron Harless and Associates, Tampa, Florida

In the psychological assessment of cortical damage, the Spiral Aftereffect Test (SAET) has become an important and controversial technique. This technique was first introduced into the literature in 1949 by Freeman and Josey. Further pioneering of the SAET was conducted by Standlee (1953) and Price and Deabler (1955).

The illusion, or aftereffect, following rotation of the Archimedes Spiral has long been known and used in experimental psychology (Boring, 1942). Research on the SAET is summarized in Table 1.

It is noted that the mechanical technique used in presenting this test has varied considerably. Some research indicates high validity in differentiating subjects (Ss) with cortical damage from normals and functional states (Price & Deabler, 1955). Some research raises question as to the usefulness of this test in diagnosing brain damage (Gilberstadt, Schein, & Rosen, 1958). Researchers have had to construct their own apparatus, and no two studies have utilized identical apparatus and presentation. These factors, as well as differences in sampling technique and criteria, may account for the variation in research results. For instance, in evaluating the apparatus over a considerable period of time, it was discovered that any deviation of the motor shaft caused ambiguous aftereffect perception.

The importance of base rates in predicting valid measurements of cortical damage using

the SAET has been reported in recent literature (Stilson, Gynther, & Gertz, 1957; Gilberstadt et al., 1958).

Berger, Everson, Rutledge, and Koskoff (1958) conducted an elaborate study using extensive clinical criteria. Results indicated negative correlation when the SAET was compared with EEG records, pneumoencephalograms, and skull X-rays. Three studies are reported using children as the experimental sample (Spivack & Levine, 1957; Davids, Goldenberg, & Laufer, 1957; Harding, Glassman, & Helz, 1957). The results seem to indicate the usefulness of the SAET in the measurement of cortical involvement in these children.

Of 16 studies reported, 10 have used four trials, and the remainder, from three to six trials in presenting the revolving spiral. Yates (1954), in his evaluation of research measuring cortical damage, indicates this area has been studied most inadequately. He refers to such criticism as: lack of comparable groups for validation studies, classification of cortical damage, the control of relevant factors such as age, sex, intelligence, subjective scoring systems, and adequate control of the testing procedures.

The present study gave particular attention to the above criticisms.

Procedure

Subjects

Four hundred and twenty children between the ages of 5 and 16 years were examined during a period of two-and-a-half years at an outpatient psychological clinic. The variety of complaints and disorders ranged from requests for a general psychological checkup to referrals for differential diagnosis

¹ The authors wish to acknowledge the considerable part played in this study by Arturo G. Gonzalez, whose skill and interest made the detailed criterion data possible.

² Now in private practice, Tampa, Florida.

Table 1

Showing the Various Mechanical Features, Experimental Procedures, Predictors, Criteria, and Results in Spiral Aftereffect Test (SAET) Research

Author(s)	Date	Degree of Spiral	rpm	Disc Size (inches)	Trials	Subjects ^a	Criterion	Results
Freeman and Josey	1949	1260	100	6	4	50 Normals 85 Psychotics	Hospital diagnosis	1. Memory impairment and no aftereffect correlated 2. Test-retest reliability .94
Standlee	1953	No details of apparatus				16 Normals 25 Psychotics	Hospital diagnosis	1. Electroshock therapy does not impair capacity to perceive aftereffect 2. No relation between perception of aftereffect and memory
Price and Deabler	1955	920	100	6	4	40 Normals 40 Psychiatric 120 Organics	1. Hospital diagnosis 2. Personal history	1. Nonorganic psychotics and normals perceived aftereffect 2. Cortical involvement cases unable to perceive aftereffect
Gallese	1956	920	90	6	4	30 Normals 41 Schizophrenics 47 Organics (CNS) 50 Organic—convulsive 12 Lobotomized Schizophrenics	Hospital diagnosis	1. SAET differentiated organic from nonorganic patients 2. Some organics more sensitive to aftereffect than others 3. Lobotomized patients perceived aftereffect 4. Perception of aftereffect unrelated to age, sex, length of hospitalization 5. Test-retest reliability (.84)
Spivack and Levine	1957	920	78	No details	4	32 Brain injured adolescent boys 35 Emotionally disturbed adolescent boys	APA Nomenclature	Emotionally disturbed adolescents perceived aftereffect more frequently than brain damaged adolescents
Page, Rakita, Kaplan, and Smith	1957	920	100	8	3	20 organic cases 20 Psychiatric cases	Hospital diagnosis	Organic patients less likely to perceive aftereffect than functional cases
Davids, Goldenberg, and Laufer	1957	920	78	8	6	24 Normal children 29 Psychiatrically disturbed children 15 Organic children	Medical and psychological evaluation	The SAET as well as the Trail Making Test possess promise as methods for assessing cortical impairment
Stilson, Gynther, and Gertz	1957	No apparatus Critical review of previous research						1. Stresses importance of the base rate in determining discrimination of any diagnostic instrument 2. SAET can be used with considerable confidence to differentiate nonorganics
Garret, Price, and Deabler	1957	920	100	6	4	40 Organics 30 Normals	Hospital diagnosis	1. SAET falls into category of phenomenological movement (phi phenomena) 2. SAET with the Graham Kendall Test form a valid battery for indicating cortical damage

Table 1—Continued

Author(s)	Date	Degree of Spiral	rpm	Disc Size (inches)	Trials	Subjects ^a	Criterion	Results
Harding, Glassman, and Helz	1957	920	78	6	4	81 Normal children	Staff decision	1. Normal children beginning at 4.5 to 6 years of age perceive aftereffect 2. No children below 55 months perceived aftereffect 3. Mental age had more direct relationship to perception of aftereffect than chronological age
Goldberg and Smith	1958	920	78	10	4	30 No pathology 24 Mixed neurologic 38 Psychiatric	Hospital diagnosis	1. All normals perceived aftereffect 2. Results warn against indiscriminate use of SAET for differential diagnosis
Aaronson	1958	920	78	6	4	65 Epileptics	Hospital diagnosis	1. Poor performance on SAET associated with impairment of visual function 2. SAET seems most sensitive to temporal lobe involvement
London and Bryan	1958	No details of apparatus				22 Normals 44 Brain injured	Hospital diagnosis	1. Structured instructions aided organics to see aftereffect as frequently as normals 2. Organics unable to see aftereffect with neutral instructions
Gilberstadt, Schein, and Rosen	1958	920	100	No details	4	87 Psychiatric 140 Neurologic	Hospital diagnosis	1. All Ss considered, the SAET diagnoses brain damage no better than base rates 2. Base rates can be helpful in determining cutting score in particular populations using the SAET as a screening technique 3. The SAET has limited usefulness
Berger, Everson, Rutledge, and Koskoff	1958	920	78	6	4	110 Neurologic cases	1. Hospital diagnosis 2. EEG 3. Pneumoencephalogram 4. Snellan chart 5. Skull x-rays 6. Visual field	1. Visual acuity correlated with perception of aftereffect 2. The aftereffect phenomena involves radial movement and Gamma movement 3. SAET performance not related to electroencephalogram, pneumoencephalogram of skull X-rays 4. SAET performance related to spinal fluid studies, visual field studies and likelihood of brain damage
Spivack and Levine	1958	No details of apparatus			Staggered trials	24 Organics 20 Normals	Staff diagnosis	1. Performance on SAET differentiates between brain damaged and normal group with slight efficiency 2. The duration of the aftereffect may be a fruitful area for research 3. Lack of perception of aftereffect by organic cases may relate to well known perseveration phenomena in brain injury cases

^a As designated in the original research.

of what was tentatively described as childhood schizophrenia. All of these Ss were given various psychological tests, but in each case a SAET of eight trials was administered. Figure 1 presents graphically the distribution of SAET scores for the cases.

Experimental group. Fifty of the children shown in Fig. 1 were selected as the "predicted abnormal" group. These children did not perceive the aftereffect phenomenon on eight trials of the SAET. Four of the original sample were dropped from the study because EEG evaluation was not obtainable. This resulted in an experimental group of 46 cases. The mean age in months for the predicted abnormal group was 122.98. There were 34 males and 12 females included in the group. The average grade placement was at the 4.0 grade level. Included in the group were 37 cases where one of the initial complaints was poor reading ability, 13 cases where speech defect was part of the presenting picture, 20 cases where the child was overly aggressive, and 22 cases where the child was said to be withdrawn. Total complaints exceed 46 since several symptoms were frequently described for a single case.

Control group. A matched control group was selected for this study. Twenty cases, matched for age, were selected from the sample shown in Fig. 1. All Ss in this group achieved a full score of eight on the SAET. Mean age for this group was 119.65 months, and the group included 11 males and 9 females. The average grade placement was at the 4.4 grade level. Within this group were two children with severe reading problems, one with speech difficulty, three overly aggressive children, and one child who was said to be withdrawn. The remainder of the group were children who were seen for routine psychological evaluation. This group was designated the "predicted normal group."

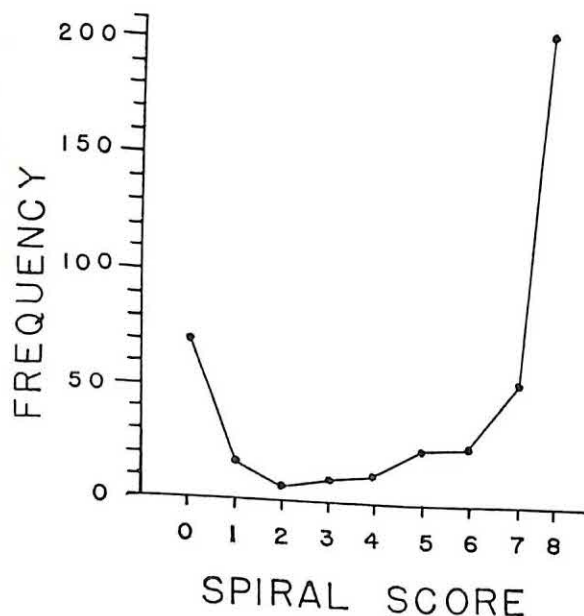


Fig. 1. Distribution of Spiral Aftereffect test scores for 420 children referred for outpatient psychological services.

Table 2
Matching Data for the Experimental
Group and Control Group

Variable	Control Group (Predicted Normal EEG)	Experi- mental Group (Predicted Abnormal EEG)
N	N = 20	N = 46
Mean age in months	119.65	122.98
Sex		
♂ Males	11	24
♀ Females	9	12
School placement, grade	4.4	4.0

Table 2 presents the data for the experimental group and the control group.

Predictor Battery

Although the primary purpose of this study was the evaluation of the SAET, it was decided to include some standard psychological instruments said to be related to the presence or absence of brain damage.

The psychological tests selected for this study are viewed as screening devices rather than diagnostic tools. It was hypothesized that psychological measurements can discriminate children of essentially normal cortical function from children having disrupted cortical function.

The Spiral Aftereffect Test. The stimulus for this procedure was an Archimedes Spiral (920°), imprinted on a 7-in. white disc, presented to Ss with a rotation speed of 80 rpm (5 rpm variability). The apparatus was battery operated and reversible (clockwise and counterclockwise rotation).²

Each S was seated approximately eight ft. from the spiral. The following instructions were given:

This is a test of the way you see. Please keep your eyes on the silver button in the center of the wheel. In a moment I will turn on the machine. The wheel will go around and I will ask you what you see. Please keep your eyes on the silver button even when I ask you questions. Do you understand?

At this point the machine was turned on. After 10 sec., as timed with a stop watch, the S was asked: "What is happening to the black line?" In most cases the S responded with the appropriate impression. If a vague response was given, the S was asked: "Is the black line moving toward the center or away

² Details of the apparatus are available from the Psychological Research and Development Corporation, Tampa, Florida.

from the center?" In all cases, Ss were able to respond to this inquiry. At the end of 20 sec. the S was told: "In a moment I will turn the machine off. Please watch the wheel very carefully when I turn the machine off." At the end of 30 sec. the apparatus was stopped and the S was asked: "What is *now* happening to the black line?" Any reported aftereffect of opposite direction from the original effect was given a score of 1. A zero score was given if no aftereffect was reported.

Eight trials were presented in an ABBABBAA order.

Reliability of this test is well known (Freeman & Josey, 1949; Gallese, 1956; Price & Deabler, 1955) with test-retest and interjudge reliability having been found to be between .84 and .94 by various experimenters.

The Bender Visual Motor Gestalt Test. This well-known test has been described elsewhere (Bender, 1938). Although commonly evaluated by clinical judgment, a five-point rating scale was developed for this study. Standard instructions were followed in the administration of this test.

The five-point rating scale was applied as an "inspection technique." The judges examined the entire Bender record of each S and noted such discrepancies as distortion, collision (one figure running into another) and rotation (a figure turned from the normal axis). Thus the test productions were rated in terms of their apparent equivalence to the original cards. Each S was given a score ranging between 0 and 4, with 0 representing what would be considered an essentially perfect production, with consideration for the age of the S (Bender, 1946).

Two judges, each with several hours' training on this technique, rated all of the cases in the study. Applying a raw data formula, a Pearson correlation of .77 was obtained. Thus, the reliability coefficient is significant well beyond the .01 level. A cutoff point was established on the rating scale whereby scores of 0, 1, and 2 were considered to be normal and scores of 3 and 4 were considered to be abnormal. In comparing the interjudge reliability using this method, agreement was obtained in 93% of the cases.

Draw-A-Person Test. This was the standard clinical test in which the S was asked to "draw a person." No further directions were given other than to encourage the S to complete the drawing.

Each drawing was rated on the single variable of "tipping." A perpendicular figure was considered to have a score of 0, while a "floating" figure, lying horizontal on the paper, was given a score of 4. Scores of 1 through 3 were assigned to various degrees of rotation between the perpendicular and the horizontal.

Two judges rated all of the drawings in the study. Applying the raw data formula to the judges' ratings, a Pearson correlation of .75 was obtained. This is significant beyond the .01 level. A second reliability evaluation was done by setting a cutoff score between 2 and 3 on the five-point scale. Thus, scores of 0, 1, and 2 were considered to be predictive of normal cortical function. Scores of 3 and 4 were

considered to be predictive of abnormal cortical function. Agreement was found on 85% of the ratings using this method.

The Wechsler Scales. For the purposes of this study, several subtests of the Wechsler Intelligence Scale for Children (WISC) were administered to all of the Ss. The scales selected were, Digit Symbol, Block Design, Arithmetic, and Digit Span. An additional Digits-Reversed Score was also included. These scales were administered and scored according to standard procedures. Except for the Digits-Reversed variable, weighted scores were used.

Criterion

Whereas all previous studies have dealt with known brain damaged groups or assumed non-brain-damaged groups, this study deals with a group of children who were supposedly non-brain-damaged, as a result of physical examination before coming to the outpatient clinic in which they were tested. Empirical evidence indicates that about 10% of clinic patients, even though physical examination is negative, have some cortical disruption which might be described as subclinical intracranial pathology. It was deemed necessary that this study would have an objective neurologic evaluation, and the EEG was chosen as the preferred criterion measure. All members of the experimental and the control groups were referred for EEG evaluation. A two-hour EEG record was taken on each child. A standard eight-channel Medcraft instrument, CS7C Model D, was used. Tracings were obtained in all leads, and a portion of the recording was a sleep record.

Only one judge was used in this experiment, since in the locality of the study there was a neuropsychiatrist available who had extensive experience in EEG with children. Each child was rated by the neurologist on a five-point scale ranging from 0 to 4, on the factor of focal findings vs. no focal findings. A second rating scale was administered to measure diffuse findings vs. no diffuse findings. In determining the final criterion, those cases which showed any EEG anomalies were called the "abnormal EEG group." Those Ss whose EEGs resulted in no findings were considered to have normal cortical function.

Six months after the original study, the EEG tracings of 18 of the Ss, including both "predicted normal" and "predicted abnormal," were selected using a table of random numbers. These records were given to the original rater for blind re-evaluation. He was asked to rate each of these records in terms of findings vs. no findings. When these ratings were finished, it was found that all 18 cases had been rerated exactly as the original ratings. Since an intrajudge correlation of 1.00 seemed highly unlikely, the results were questioned. In order to insure a more careful evaluation of the reliability of the EEG ratings, 36 cases were selected for re-evaluation. Half of these were predicted normal and half were predicted abnormal. In all cases, identifying data were removed and the neuropsychiatrist rating the cases re-evaluated the records with no idea of which were normal and which were abnormal. Between 6 and 18 months

Table 3

Correlation of Various Predictors with EEG Criterion and Intercorrelation with Each Other

Predictor Variable	EEG	Spiral	Bender	DAP
EEG (criterion)	.77 ^{a,*}			
Spiral Aftereffect	.74 ^{a,*}	.84*		
Bender Gestalt	.36 ^{a,*}	.42 ^{a,*}	.77*	
Draw-A-Person	.30 ^{a,*}	.35 ^{a,*}	.52 ^{a,*}	.75*
WISC				
Digit Symbol	.25	.37*	.36*	.09
Block Design	.37*	.54*	.45*	.19
Arithmetic	.41*	.43*	.32	.00
Digit Span	.28	.30	.31	.02
Digits Reversed	.08	.01	.23	.10

^a Phi coefficient.

* Significant at .01% level or better (.01 level = .345; .05 level = .166).

had elapsed between the original ratings and this final rerating.

The same criteria of findings vs. no findings was applied, and the resulting data yielded a phi coefficient of .77. The neuropsychiatrist was able to re-rate the EEG tracings reliably in 87% of the cases. This sort of reliability is unusual in view of what has been found in the literature. It should be stated, however, that the EEG records taken on these children were more carefully done than one might expect in the course of a usual EEG evaluation. The judge who rated this data had long experience in electroencephalography with children. A good deal of time was spent on the inspection of the records. It was determined that it would be impossible to evaluate predictors adequately if the criteria itself were unstable.

Results

The Spiral Aftereffect Test

All Ss were given eight trials on the SAET, and scores of 6, 7, and 8 were considered to be predicted normal. A score of 0 through 5 was considered to be predicted abnormal. This dichotomy was compared with the EEG ratings of findings vs. no findings.

The Bender Visual-Motor Control Test

Scores on the five-point scale of the Bender were separated into predicted normal (score of 0 or 1) and predicted abnormal (score of 2, 3, or 4), and compared with the EEG criteria.

The Draw-A-Person Test

Scores on the five-point scale of the Draw-A-Person test were dichotomized: predicted

normal (score of 0 or 1), and predicted abnormal (score of 2, 3, or 4), and compared with the EEG criteria.

The Wechsler Scales

The raw scores on the Wechsler Scales were compared with the EEG criterion of findings vs. no findings using the point biserial correlation technique.

Intercorrelations were calculated among the various predictor variables. Table 3 presents the correlations among the various predictors and the EEG criterion. Unless otherwise indicated, point biserial correlation technique was used.

It will be noted on Table 3 that the SAET had the highest correlation with the EEG criterion. The Bender, the Draw-A-Person test, the WISC Arithmetic, and the WISC Block Design all correlated beyond the .01 level of significance with the EEG criteria.

Significant intercorrelations were found between the SAET and the Bender, between the SAET and the Draw-A-Person test, between the SAET and the WISC Digit Symbol, and between the SAET and the WISC Arithmetic.

The Bender correlated significantly, with the SAET, the Draw-A-Person test, the WISC Digit Symbol, and the WISC Block Design.

The Draw-A-Person test correlated significantly with the SAET and the Bender.

Table 4 presents the percentage of "hits" in predicting normal EEG records and abnormal

Table 4

Accuracy of Prediction of Normal and Abnormal EEG Records by the Various Predictor Variable, in Percentage Hits

Predictor Variable	Percentage of Normal EEG Records Correctly Predicted (Hits)	Percentage of Abnormal EEG Records Correctly Predicted (Hits)
Spiral Aftereffect test	1.00	.86
Bender Gestalt	.65	.70
Draw-A-Person	.60	.77
Digit Symbol	.58	.73
Block Design	.67	.71
Arithmetic	.54	.71
Digit Span	.37	.81
Digits Reversed	.37	.72

mal EEG records for all the predictor variables.

An attempt was made to analyze the characteristics of the EEG records. This was found to be difficult in that the different records were described with such a variety of terminology that statistical analysis was questionable. Eight characteristics of the records were chosen in order to present some definitive data.³ Table 5 presents this brief analysis of the EEG records.

Discussion

The data presented above would seem to indicate that the SAET and other psychological tests are variously effective for discriminating children exhibiting abnormal electroencephalographic records from children who exhibit normal records. These findings support some of the earlier work done with the SAET. The SAET seemed to be particularly effective as a discriminating technique. The sample was considered normal by general physical examination. This raises the hypothesis that psychological tests are perhaps more sensitive indicators of minimal cortical disruption than the usual medical-neurological examinations.

The present study differs in its findings from some of the previous research on the SAET, particularly one study which used EEG as the criterion. The EEG record must be considered an *intermediate* criterion rather than an *ultimate* criterion. There are a variety of ways and means by which the EEG may be used as a validating instrument. In the present study, it was used with the considerable care, and the reliability data presented lends some support to this technique as an adequate criterion. The fact that it is not an ultimate criterion is evidenced by the fact that one of the children whose EEG record was rated "no findings" had a grandmal seizure in the clinic. It is conceivable, in view of suspected criterion fallibility and less than perfect intrajudge EEG reliability, that the predictability of the SAET as well as other psychological measurements may actually be better than the present data indicates. Since,

³ Analysis of the EEG records continues and will be presented in a later paper.

Table 5

A Brief Analysis of Some Characteristics of the EEG Records of the Predicted Normal and the Predicted Abnormal Groups

	Predicted Normal Group	Predicted Abnormal Group
Mean occipital alpha rhythm (per second)	8.7	8.2
Mean rating of paroxysmal activity (rated 1-4)	1.08	2.06
Mean rating of build-up under hyperventilation (rated 1-4)	2.0	2.31
Mean rating of dysrhythmia or disorganization (rated 1-4)	2.25	2.52
Percentage of records with focal slow activity	.17	.25
Percentage of records with spikes		
Single	0	.40
Multiple	0	.18
Spike & Dome	0	.14

however, we must be dependent upon an intermediate criterion, all we can do is attempt to establish the criterion used with the greatest of care and attention to questions of reliability and validity.

The most significant question in the evaluation of the SAET, as well as any other diagnostic technique, is the question of base rate. This matter has been previously discussed in the introduction to this study. It was decided not to apply the base rate formula to the present data since there is no information available of random samplings of unselected EEG tracings in children between the ages of 5 and 16 years. This would seem to be a very necessary area for further research.

In judging the effectiveness of any instrument, one must consider its application. Psychological tests, when considered in terms of the presence or absence of brain damage, are *screening* devices. In and of themselves they cannot be diagnostic. The diagnosis of the presence of brain damage would be a medical problem, whether it be through laboratory tests, histological examination, or results of treatment. In the present study, the

SAET was the best of a series of psychological tests for the prediction of EEG findings. Other psychological tests were significant but less accurate in their predictability. Undoubtedly we will always be dealing with the question of false positives and false negatives. It is still a moot question whether errors in prediction are based on inaccuracies of the predictor variables, or in the criterion itself. Much careful research is necessary to clarify these points. For the time being, however, since clinical psychologists are called upon to give tests and, as part of their testing procedure, to make some determination of who shall be referred to the neurologist and who shall not be referred, the SAET, probably in combination with one or more standard psychological instruments, would seem to be the best available technique for this purpose.

REFERENCES

- AARONSON, B. S. Age, intelligence, aphasia, and the spiral aftereffect in an epileptic population. *J. clin. Psychol.*, 1958, 14, 18-21.
- BENDER, L. *A visual motor gestalt test and its clinical use*. New York: American Orthopsychiat. Assn., 1938. Pp. 176.
- BENDER, L. *Instructions for the use of visual motor gestalt test*. New York: American Orthopsychiat. Assn., 1946. Pp. 4-5.
- BERGER, D., EVERSON, R., RUTLEDGE, L., & KOSKOFF, Y. D. The spiral aftereffect in a neurological setting. *J. consult. Psychol.*, 1958, 22, 249-255.
- BORING, E. G. *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century, 1942. Pp. 591-592.
- DAVIDS, A., GOLDENBERG, L., & LAUFER, M. W. The relation of the Archimedes spiral aftereffect and the trail making test to brain damage in children. *J. consult. Psychol.*, 1957, 21, 429-433.
- FREEMAN, E., & JOSEY, W. E. Quantitative visual index to memory impairment. *AMA Arch. Neurol. Psychiat.*, 1949, 62, 794-797.
- GALLESE, A. J., JR. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, 12, 254-258.
- GARRETT, E. S., PRICE, A. C., & DEABLER, H. L. Diagnostic testing for cortical brain impairment. *AMA Arch. Neurol. Psychiat.*, 1957, 77, 223-225.
- GILBERSTADT, H., SCHEIN, J. D., & ROSEN, A. Further evaluation of the Archimedes spiral aftereffect. *J. consult. Psychol.*, 1958, 22, 243-248.
- GOLDBERG, L. R., & SMITH, P. A. The clinical usefulness of the Archimedes spiral in the diagnosis of organic brain damage. *J. consult. Psychol.*, 1958, 22, 153-157.
- HARDING, G. F., GLASSMAN, S., & HELZ, W. C. Maturation and the spiral aftereffect. *J. abnorm. soc. Psychol.*, 1957, 54, 276-277.
- LONDON, P., & BRYAN, J. H. The influence of instructions on spiral aftereffect reports. *Amer. Psychologist*, 1958, 13, 7.
- PAGE, H. A., RAKITA, G., KAPLAN, H. K., & SMITH, N. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, 21, 89-91.
- PRICE, C. A., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, 19, 299-302.
- SPIVACK, G., & LEVINE, M. The spiral aftereffect and reversible figures as measures of brain damage and memory. *J. Pers.*, 1957, 25, 767-778. (Mimeo.)
- SPIVACK, G., & LEVINE, M. Spiral aftereffect, parameters and correlates. Unpublished manuscript, Devereux Foundation for Res. & Trng., 1958.
- STANDLEE, L. S. The Archimedes negative aftereffect as an indication of memory impairment. *J. consult. Psychol.*, 1953, 17, 317.
- STILSON, D. W., GYNTHIER, M. D., & GERTZ, B. Base rate and the Archimedes spiral illusion. *J. consult. Psychol.*, 1957, 21, 435-437.
- YATES, A. C. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, 51, 359-379.

(Received January 19, 1959)

A COMPARISON OF SHADING RESPONSES OBTAINED WITH TWO RORSCHACH METHODOLOGIES FROM PSYCHIATRIC AND NON- PSYCHIATRIC SUBJECTS¹

PATRICIA F. WALLER²
University of North Carolina

Baughman has developed a new inquiry procedure for the Rorschach Ink Blot Test (Baughman: 1954, 1958). This procedure utilizes a paired comparison presentation of a series of modifications of the original blot. For example, if a subject (S) reported a flower garden on Card X, instead of the usual questioning on the inquiry, the S is presented the original card and an identical card in which the color has been eliminated. He is then asked if he can still see the flower garden in the achromatic card, in which card is it clearer, and why. The original card is then removed and replaced by a card in which the shading has been eliminated, and the questioning proceeds as before. The cards continue until there is only an outline of the blot remaining, but it is rarely necessary to continue questioning throughout the entire series.

It has been shown that this paired comparison inquiry procedure elicits from non-psychiatric Ss many more shading responses than does the regular inquiry procedure (Baughman, in press; Marimon, in press). It appears that the questions involved in the paired comparison inquiry are easier to answer than those in the conventional inquiry. Rather than asking, "What is it about the blot that makes it look like a bat?" one asks, "Do you still see the bat here? Is it as clear? What makes it clearer over here?" Although

this technique takes slightly longer, it appears that it does not demand so much from the S. The questions do not seem so difficult to answer.

Method

Problem

If shading is truly related to anxiety, as has often been hypothesized (Beck: 1944, 1951; Binder, 1937; Hertz, 1940; Klopfer, 1937; Rorschach & Oberholzer, 1942), then it would seem reasonable to assume that shading responses would be more prominent in a psychiatric population than in a nonpsychiatric population. However, if it is also assumed that psychiatric patients are under greater stress, then it may be reasonable to assume that it would be more difficult for them to identify the stimuli giving rise to their behavior. Although they may be reacting to shading, the stress they are under may make it more difficult for them to report it. Cox and Sarason (1954) have made this point and have actually found fewer texture responses in high anxiety Ss than in low anxiety Ss. An easier, more objective inquiry may enable such Ss to express the shading determinant whenever they are using it. On the basis of this reasoning, it was predicted that psychiatric patients would show a greater difference in shading responses between the regular and the Baughman inquiries than would nonpsychiatric Ss.

Subjects

Data from a previous study by Baughman (1959) provided information from a nonpsychiatric population.³ Ss were 162 employees of three insurance companies in a southern city. The psychiatric Ss were 60 patients from the inpatient and outpatient services in a hospital in a nearby town. Psychiatric Ss were selected according to three criteria. First, they were to be showing no obvious signs of psychosis. Second, there should be no question of organic impairment, and third, they should be literate.

¹ This article is based upon a doctoral dissertation completed at the University of North Carolina, August, 1958. This paper was presented at the 1958 meeting of the American Psychological Association, Washington, D. C.

² Now at R. B. Jackson Memorial Laboratory, Bar Harbor, Maine.

³ The author is indebted to Emmett Earl Baughman, who made available data from his research.

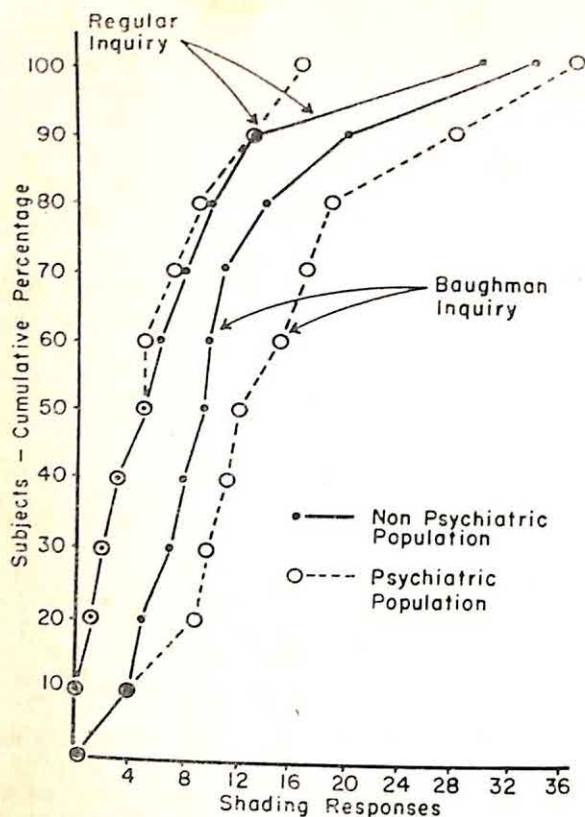


Fig. 1. Cumulative distributions of shading responses obtained from psychiatric and nonpsychiatric populations, using two Rorschach methodologies.

Whether a patient was considered psychotic was decided by the psychiatrist in charge. In some cases, later evaluation led to a psychotic diagnosis, but in such instances the diagnosis could be termed debatable, and the *S* was retained in the study. The psychiatric and the nonpsychiatric groups were comparable in terms of mean number of responses to the Rorschach (20.10 and 22.05, respectively) and median education (high school graduate).

Procedure

The nonpsychiatric population consisted of two groups of 81 *Ss* each, one group receiving the regular inquiry and the other the paired comparison. The psychiatric population consisted of two groups of 30 *Ss* each, one group receiving the regular inquiry and the other the paired comparison. The same examiner tested all the psychiatric *Ss* and many of the nonpsychiatric, so that the method of administration was comparable. All records were scored by two scorers. Although scoring included the various forms of *Y*, *V*, and *T*, interscorer reliability was based on the sum of shading responses for each record, that is, the total number of responses involving shading of any kind. Interscorer reliability (Pearson *r*) ranged from .90 to .98.

Results

A comparison of the groups on the basis of the sum of shading responses is shown in Fig. 1. Because of the unequal *Ns* for the two groups, the results are depicted in cumulative percentages. Table 1 lists the means and standard deviations of number of shading responses for the two inquiries with the psychiatric and the nonpsychiatric populations. On the basis of the sum of shading responses, both the psychiatric and the nonpsychiatric groups show significantly more shading on the Baughman inquiry procedure ($p < .01$). However, the psychiatric population shows even more than does the nonpsychiatric population. A *t* test of the difference between the differences was significant at $< .01$ level, confirming the hypothesis that the psychiatric population would show a greater difference between the regular and the paired comparison inquiries on the basis of the shading response.

The relative efficiency of the two inquiry procedures for discriminating between psychiatric and nonpsychiatric populations was tested. As can be seen in Fig. 1, the regular inquiry shows virtually no difference between the two populations, and where the differences occur, they are the result of the nonpsychiatric population giving more shading than the psychiatric population. The direction of difference supports the theory of Cox and Sarason but is contrary to expectations based on orthodox Rorschach theory. However, the differences between the groups on the basis of the regular inquiry are not significant ($t = .69$, 109 *df*). On the basis of the Baughman paired comparison inquiry, the

Table 1
Shading Responses for Psychiatric and Nonpsychiatric Populations on Each of Two Inquiry Procedures

Inquiry	Psychiatric	Nonpsychiatric
Regular	$\bar{X} = 5.5$	$\bar{X} = 6.3$
	$SD = 4.7$	$SD = 6.2$
	$N = 30$	$N = 81$
Baughman	$\bar{X} = 14.8$	$\bar{X} = 10.6$
	$SD = 8.6$	$SD = 6.4$
	$N = 30$	$N = 81$

psychiatric population gives more shading, as would be predicted from Rorschach theory. A *t* test of the differences was significant at $<.02$.

Discussion

The hypothesis that psychiatric Ss would show a greater difference between the regular and the paired comparison inquiries in regard to the shading response was supported by the results. Furthermore, the differences found indicated that the paired comparison inquiry differentiates between psychiatric and nonpsychiatric populations significantly better than does the regular inquiry. These results lend support to the assumption that psychiatric patients are more sensitive to the shading determinant but have more difficulty reporting it. In other words, the regular inquiry may not differentiate between Ss who give little shading because they are not utilizing it and Ss who give little shading because they are having difficulty determining the basis of their response. The Baughman paired comparison inquiry procedure makes it easier for both psychiatric and nonpsychiatric Ss to report the determinants involved in their responses. The fact that on the regular inquiry procedure, psychiatric and nonpsychiatric Ss give essentially the same amount of shading, and on the easier Baughman inquiry psychiatric Ss give significantly more shading than nonpsychiatric Ss, lends support to the common contention that shading is in some way related to anxiety and should therefore be more prominent in a psychiatric population.

Summary

Sixty psychiatric and 162 nonpsychiatric Ss were compared on the basis of Rorschach shading responses obtained with two inquiry procedures, the regular and the Baughman paired comparison. It was predicted that the psychiatric group would show a greater dif-

ference between the two inquiry procedures, since they should be responding more to shading but have greater difficulty reporting it, because of the stress they would presumably be under. The results supported this hypothesis at $<.01$ level of significance. The results also indicated that on the basis of total number of shading responses, the regular inquiry procedure failed to differentiate between the psychiatric and the nonpsychiatric groups, whereas the Baughman paired comparison inquiry differentiated at $<.02$ level of significance.

REFERENCES

- BAUGHMAN, E. E. A comparative analysis of Rorschach forms with altered stimulus characteristics. *J. proj. Tech.*, 1954, 18, 151-164.
- BAUGHMAN, E. E. A new method of Rorschach inquiry. *J. proj. Tech.*, 1958, 22, 381-389.
- BAUGHMAN, E. E. The effect of inquiry method on Rorschach color and shading scores. *J. proj. Tech.*, 1959, 23, 3-7.
- BECK, S. J. *Rorschach's test: I. Basic processes*. New York: Grune & Stratton, 1944.
- BECK, S. J. The Rorschach test: A multi-dimensional test of personality. In H. H. Anderson & Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951.
- BINDER, H. The "light-dark" interpretation in Rorschach's experiment. *Rorschach Res. Exch.*, 1937, 2, 37-42.
- COX, F. N., & SARASON, S. B. Test anxiety and Rorschach performance. *J. abnorm. soc. Psychol.*, 1954, 49, 371-377.
- HERTZ, MARGUERITE R. The shading response in the Rorschach ink blot test: A review of its scoring and interpretation. *J. gen. Psychol.*, 1940, 23, 123-167.
- KLOPPER, B. The shading responses. *Rorschach Res. Exch.*, 1937-38, 2, 76-79.
- MARIMON, NANCY F. A new-type Rorschach inquiry used with children. Unpublished master's thesis, Univer. of North Carolina.
- RORSCHACH, H., & OBERHOLZER, H. The application of the form interpretation test. In H. Rorschach (Ed.), *Psychodiagnostics* (Trans. by P. Lemkau and B. Kronenbourg). Bern: Verlag Hans Huber, 1942.

(Received February 2, 1959)

THE RELATIONSHIP BETWEEN KLOPFER'S RORSCHACH PROGNOSTIC RATING SCALE AND PHILLIPS' CASE HISTORY PROGNOSTIC RATING SCALE¹

CLAUDENE SEIDEL

George Washington University

Clinical use of the Rorschach test has led to the observation that there is often an extreme discrepancy between the general nosological picture that a patient presents overtly and symptomatically, and the ego structure as revealed by the Rorschach. Some of the psychological implications of this observation, according to Klopfer, are that the Rorschach record usually reflects the individual's ego organization as it is shaped by his pre-morbid level of maturity as well as his ego defenses and character formation, and that the ego organization may not always be easily recognizable in overt symptoms (Klopfer, Ainsworth, Klopfer, & Holt, 1954).

The maturity of the male patient may also be reflected in his history of sexual adjustment and marital responsibilities as well as in his relationships to parents and peers from childhood to adulthood. The psychiatric patient's premorbid maturity may not be readily apparent from his observable behavior at the time of admission to the hospital. Scales have been developed to deal with the patient's underlying ego organization and premorbid social and sexual history.

Phillips (1953) has developed a rating scale which utilizes case history data for predicting the outcome of shock treatment in schizo-

phrenia. This scale includes three areas of case history data: the premorbid history (sexual and social adjustment), possible precipitating factors, and signs of the disorder (affect and thought). Phillips found that, in general, case history items pertaining to the recent premorbid past (found in Sec. I of the scale) showed the clearest relationship to outcome with his group of patients. These items are: recent sexual adjustment, social aspects of the recent sexual life, past history of personal relations, and the recent adjustment in personal relations.

It was thought that Phillips' scale and Klopfer's Rorschach Prognostic Rating Scale (RPRS) (Klopfer, Kirkner, Wisham, & Baker, 1951) might be of value in determining whether there is a relationship between the level of maturity reflected in a schizophrenic patient's case history and his Rorschach record, and also might be useful in predicting outcome of the illness.

Method

Sample

The sample was selected from the file of regular referrals (for diagnostic evaluation) to the Psychology Branch of Saint Elizabeths Hospital, Washington, D. C. (This is a mental hospital of over 7,000 patients made up of voluntary admissions and commitments from all socioeconomic levels.) The criteria set up for the selection of the sample were:

1. Only white, male schizophrenic patients were selected because the sample on which Phillips tested his scale was comprised of males.

2. The patients in the sample group should have been given a Rorschach test within four months after their admission date. (Two patients in the sample group did not strictly meet this criterion; one was tested within five months of admission and the other within six months.)

¹ This paper is based on a thesis submitted to the faculty of the Columbian College of the George Washington University, Washington, D. C., in partial satisfaction of the requirements for the degree of Master of Arts. The author is presently a clinical psychologist with the Arlington (Virginia) Mental Hygiene Clinic. Acknowledgment is made to Albert D. Annis, formerly Director of Research in Psychology, St. Elizabeths Hospital, Washington, D. C., for his assistance in setting up the study, his advice throughout the course of the investigation, and his ratings on the Rorschach Prognostic Rating Scale.

3. The patients should have been hospitalized on or before May 1, 1954, so that an evaluation could be made of the outcome of their illness after three years (i.e., on May 1, 1957).

4. The Rorschach records used in this study were obtained before these patients had received tranquilizing drugs.

5. Only legal residents of the District of Columbia and Veterans Administration cases were included so that comparably complete case histories could be obtained for all members of the group.

6. The age range was between 20 and 50.

7. Only patients whose Rorschach records contained at least five responses were chosen for study.

In order to select 100 cases it was necessary to choose all patients conforming to these criteria that were referred for testing between February 1948 and December 1954.

The distribution of the sample group according to diagnosis (type of schizophrenia) is as follows: catatonic—33, paranoid—32, hebephrenic—5, simple—2, schizo-affective—2, undifferentiated—22, and type undetermined—4. The mean age of the sample was 31.9 years.

Rating Procedure

Phillips' Scale. The case histories were rated on Phillips' scale by the author. Each rating was made on that portion of the case history that was obtained at the psychiatric case study interview and from relatives and friends at the time of admission. Later data, pertaining to the period of hospitalization, was not included in the scale ratings.

The ratings on premorbid social and sexual history (Sec. I of Phillips' scale) were used in this study rather than the ratings on the remainder of the scale, because in Phillips' study, Sec. I showed the most significant relationship to outcome of illness. Preliminary correlations with the Klopfer ratings were found to be as high or higher when ratings on Sec. I of Phillips scale were used as when the ratings on the total scale were used. The correlation between Phillips I and Phillips total was found to be .96 for the sample group used in this study.

Each patient was rated on five subsections of Sec. I and the average of these ratings was determined. There are actually six subsections in Sec. I, but since two of these (social aspects of recent sexual life—30 years and above; social aspects of recent sexual life—below 30 years of age) are mutually exclusive, each patient could be rated on only five subsections.

Klopfer's Scale. The ratings on Klopfer's Rorschach scale were made without knowledge of the scores obtained on Phillips' scale.²

² Klopfer's method for determining scores was followed with one exception. Klopfer subtracts 1.0 from the weighted form level score when there is a discrepancy of 3.0 or more between the lowest and highest form level ratings, provided the lowest form level rating is a minus score. In this study 1.0 was

Outcome Criterion

One of the stated purposes of this study was to learn whether Phillips' scale and Klopfer's scale might be useful in predicting the outcome of illness of schizophrenic patients. Two criteria were set up to investigate this question. The first of these was admission to the hospital on or before May 1, 1954, so that an evaluation of outcome of illness could be made after a three-year period. Secondly, those patients that left the hospital prior to three years had to receive a psychiatric evaluation of Social Recovery or Recovery. It was found in checking the records that 7 patients did not strictly meet the first criterion, so they had to be excluded from the prognostic evaluation. This left 93 patients that met the first criterion. Thirty patients did not meet the second criterion because they left the hospital prior to three years with a psychiatric evaluation of Improved or Unimproved. These 30 patients are a heterogeneous group. Some of them were released against medical advice, others were transferred to another hospital, and others were discharged because they were considered to have received maximum hospital benefit. Since these 30 patients could not be analyzed properly as to outcome of illness, they also had to be excluded from the prognostic evaluation (for reference purposes, the 37 patients that did not meet the criteria are called the "other" group). This left 63 patients who met the criteria: 31 that were hospitalized continuously for three years or longer, and 32 that were discharged in less than three years with a psychiatric evaluation of Social Recovery or Recovery.

Results

Product-moment correlations were computed for the original sample group of 100 cases and the prognostic evaluation group of 63 cases. Table 1 gives the correlations between the social and sexual history ratings (Phillips I) and: (a) Klopfer's total (final) prognostic score ratings, (b) Klopfer's form level ratings, and (c) Klopfer's determinant ratings. The term "determinant rating" refers to the sum of the ratings on the first five sections of the RPRS, i.e., everything except form level.

Point biserial coefficients were computed to determine the correlation between the Klopfer ratings and the dichotomous variable of Social Recovery or Recovery in less than three years, as against continuous hospitalization

subtracted only when the lowest form level rating was -1.5 or lower. This exception was made because with psychotic subjects it is often difficult to obtain a satisfactory inquiry and therefore specifications of responses are difficult to evaluate.

Table 1

Pearson Correlations between Klopfer's RPRS Ratings and Phillips' Social and Sexual History Ratings

	Prognostic Evaluation Group (<i>N</i> = 63)	Original Sample Group (<i>N</i> = 100)
Between RPRS total and Phillips I	.30*	.24*
Between RPRS form level and Phillips I	.46**	.36**
Between RPRS determinants and Phillips I	.19	.15

* Significant at .05 level.

** Significant at .01 level.

for three years or more. Point biserial coefficients were also computed to determine the correlation between premorbid social and sexual adjustment and the same dichotomous variable. These coefficients are presented in Table 2.

Chi square values were determined to learn whether there is a relationship between the Rorschach ratings and diagnosis, and between the social and sexual history ratings and diagnosis, and between these ratings and age. The Klopfer total ratings were divided into five categories: High (scores of +2.61 and above), Medium High (+1.50 to +2.60), Medium (+.18 to +1.49), Medium Low (+.17 to -2.49), and Low (-2.50 and below). These categories were set up on the basis of Klopfer's Groups III, IV, and V (Klopfer et al., 1951). The combined High and Medium High categories correspond to Klopfer's Group III (better than 50-50 chance; any treatment will be of some help), the combined Medium and Medium Low categories to Klopfer's Group IV (50-50 chance), and the Low category to Klopfer's Group V (a difficult case that may be helped somewhat but is generally a poor treatment prospect). Since there was no comparable basis for grouping the social and sexual history ratings, they were divided into only two groups: above the median, and at the median and below. The variable of age was divided into two groups: below 30 years of age (i.e., 20-29), and 30 years of age and above (i.e., 30-50). This point of division was chosen because Phillips'

scale is divided in this manner for one sub-section of Sec. I.

The computed chi square for the relationship between the Klopfer total categories and diagnosis is 16.35, which is significant at the .05 level. The relationship between the social and sexual history ratings (Phillips I) and diagnosis shows a chi square value of 7.4, which is also significant at the .05 level. The computed chi square for the relationship between the Klopfer total groupings and age is .70, which is not sufficient to be statistically significant. Neither is the relationship between the social and sexual history ratings and age statistically significant; here the chi square value is 1.00.

Discussion

The present study represents two different approaches to the study of the schizophrenic patient. One approach stresses the importance of the past history of the individual, especially aspects of his premorbid sexual and social adjustment. The other approach is primarily concerned with the patient's present and potential ego strength, or adjustment level, as reflected by the Rorschach.

The significant correlation between the ratings on Sec. I of Phillips' scale and the total ratings on Klopfer's scale supports the hypothesis of a relationship between the social and sexual history and the ego strength of schizophrenic patients. These findings are in agreement with the related findings of Fine, Fulkerson, and Phillips (1955) with normal subjects. These investigators found that the lower the social adequacy of the individual, the higher the degree of his maladjustment.

Table 2

Point Biserial Correlations between Ratings and Social Recovery or Recovery in Less Than 3 Years/Continuous Hospitalization for at Least 3 Years

Ratings	Correlations with Outcome
Social and Sexual history (Phillips I)	.46**
Klopfer total	.40**
Klopfer form level	.44**
Klopfer determinants	.34**

** Significant at .01 level.

In the present study, the poorer the premorbid social and sexual adjustment, the lower the ego strength.

According to the results obtained in this study, the social and sexual history of the schizophrenic patient is significantly related to his reality ties (as reflected by RPRS form level ratings) but is not significantly related to aspects of ego strength other than reality testing (reflected by determinant ratings on the RPRS). It seems likely that form level, as rated on the RPRS, reflects the more direct manifestation of a psychotic breakdown in ego strength, whereas the determinant ratings of the RPRS reflect, to a greater extent, the type of ego defenses and potential emotional resources of the individual. It is likely that the Phillips case history data reflect a manifest record of poor sexual and social contacts without adequate consideration of the underlying psychodynamics of sexual and social disturbance which the determinant score would be more likely to reflect. It is therefore not surprising that a significant relationship was found between the premorbid social and sexual data and form level on the RPRS, but not between the sexual and social history and the Rorschach determinants.

The results indicate that the RPRS determinant scores as well as the RPRS form level ratings are significantly related to the outcome of the schizophrenic illness. This suggests that the underlying defenses and potential emotional resources of the individual as well as the extent of his deficiencies in reality testing are important in influencing the outcome of the illness.

The significant correlations with outcome of illness suggest that ratings on social and sexual history, Klopfer form level, Klopfer total, and Klopfer determinants would all be useful in predicting the outcome of illness of schizophrenic patients.³ The present study therefore contributes toward the validation of the RPRS by supplying further data regarding its use with schizophrenic patients. This

³ It should be pointed out that the revision of Klopfer's form level method (see fn. 2) in the present study probably slightly increased the correlations in which Klopfer form level or Klopfer total ratings are involved.

study also confirms the results obtained by Phillips.

Both the Klopfer total ratings and the social and sexual history ratings show a significant relationship to diagnosis. The distributions suggest that paranoid schizophrenics received more favorable scores on both measures than did other subclassifications of schizophrenia.

Summary

One hundred male schizophrenic patients were rated on Phillips' Prognostic Rating Scale and Klopfer's Rorschach Prognostic Rating Scale and 63 could be followed up to determine their status as either discharged, recovered, or still hospitalized.

1. Section I of Phillips' scale was significantly related to RPRS ego strength (total) and reality testing (form level).

2. No significant relationship was found between Sec. I of Phillips' scale and RPRS determinants.

3. Section I of Phillips' scale and RPRS ratings for ego strength (total), form level, and determinants were significantly related to social recovery.

4. Section I of Phillips' scale and RPRS ego strength were significantly related to diagnosis (type of schizophrenic reaction).

5. No significant relationship was found between the age of the patient and RPRS ego strength or between age and Sec. I of Phillips' scale.

6. Implications for predicting the outcome of illness of schizophrenic patients were discussed.

REFERENCES

- FINE, H. J., FULKERSON, S. C., & PHILLIPS, L. Maladjustment and social attainment. *J. abnorm. soc. Psychol.*, 1955, 50, 33-35.
- KLOPFER, B., AINSWORTH, MARY D., KLOPFER, W. G., & HOLT, R. R. *Developments in the Rorschach technique*. Vol. I. *Technique and theory*. Yonkers-on-Hudson, N. Y.: World Book, 1954.
- KLOPFER, B., KIRKNER, F. J., WISHAM, W., & BAKER, GERTRUDE. Rorschach Prognostic Rating Scale. *J. proj. Tech.*, 1951, 15, 425-428.
- PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, 117, 515-525.

(Received January 29, 1959)

THE EFFECTS OF CHEMOTHERAPY ON LENGTH OF STAY AND RATE OF RETURN FOR PSYCHI- ATRICALY HOSPITALIZED PATIENTS¹

ROBERT B. ELLSWORTH

Veterans Administration Hospital, Ft. Meade, South Dakota

AND WILLIAM H. CLAYTON

Veterans Administration Hospital, Salt Lake City, Utah

From the results of many well controlled investigations of the effects of ataractic drugs on psychiatrically hospitalized patients, there can be little doubt but that many of these drugs act as a *stimulus* in modifying the psychotic's behavior and symptomatology. As Staudt and Zubin (1957) point out, however, the important question is not so much whether the therapeutic agent stimulates an immediate change, but whether the agent has any significant effect on longitudinal recovery rates. The use of control and experimental groups in an adequate double blind design, in which the experimental (drug) group was found to have improved significantly in comparison with the control group, does not begin to answer the important question raised by Staudt and Zubin. In such studies, the use of the control group allows the researcher to reach conclusions regarding only the immediate stimulus effect of the drug under investigation.

Perhaps the best way to evaluate the overall contribution of these drugs would be to compare the release rates, length of hospitalization, and return rates of two identical hospitals; one committed to a therapeutic involvement with the psychotic patient without drugs, the other relying on drugs as the major treatment method. Another possibility would be to compare the same hospitals' discharge and return rates before and after the introduction and extensive use of drugs as a

treatment method. It is the purpose of this paper to present such comparative data in an effort to evaluate the effect of ataractic drugs on the over-all rehabilitation of the psychotic patient.

Procedure

The hospital under consideration is a small, well staffed Hahn-type Veterans Administration Hospital. The Hahn-type hospital was designed so that each ward operated as a small treatment unit (24 to 32 beds), each with its own staff, dining facilities, and day rooms (Anderson & Fechner, 1956). Until very recently, each treatment ward in this hospital was staffed by a treatment team consisting of a psychiatrist, psychologist, social worker, and nurse. Only rarely did a treatment team have responsibility for more than two treatment units. The aide-patient ratio ranged from 1:2 for closed wards to 1:8 for some open wards (all 3 shifts). By the minimum suggested standards for NP hospital staffing ratios (American Psychiatric Association), this hospital was considered to be very adequately staffed.

The two periods used to evaluate the effectiveness of the widespread use of ataractic drugs in this well staffed hospital were: July 1, 1953 to July 11, 1954 (1 to 2 years prior to the introduction of ataractic drugs to the hospital); and July 1, 1956 to April 29, 1957. A recent survey revealed that 89% of all psychotic patients were on ataractic medication continuously from the time of their admission up to (and often beyond) the time of their discharge, and that 96% of all psychotics received ataractic medication sometime during the course of their hospitalization. Thus, the major change in this hospital between 1953-54 and 1956-57 was the introduction and widespread use of ataractic drugs in the treatment of the psychotic patient.

The patients under consideration were the 200 consecutive psychotic admissions between July 1, 1953 and July 11, 1954; and the 200 consecutive psychotic admissions between July 1, 1956 and April 29, 1957. All patients were male, nongeriatric veterans who

¹ The data for this study were gathered at the VA Hospital, Salt Lake City, Utah, under the support of the Psychiatric Evaluation Project, Richard L. Jenkins, Director.

Table 1
Length of Hospitalization and Return Rates of 1953-54 (predrug)
Psychotic Admissions and 1956-57 (drug) Admissions

	1953-54 Predrug	1956-57 Drug	Significance of Difference*
Median length of hospitalization of all patients ($N = 200$)	116 days	110 days	N.S.
Return rate to hospital within 1 year of discharge of all patients ($N = 150$)	25.3%	31.3%	N.S.
Median length of hospitalization of 134 1953-54 patients and 134 1956-57 patients matched on marital status and residence	112 days	92 days	N.S.

* Median test, chi square analysis.

were diagnosed as functionally psychotic at the time of admission. The admission diagnosis of functional psychosis was used since an admission diagnosis of psychotic was 86% accurate as compared with the re-discharge diagnosis. Actually, the diagnosis of a re-covered patient, who was clearly psychotic at the time of admission, was often changed to one of non-psychosis upon discharge. The admission diagnosis thus appeared to be adequate for the purposes of the study. Because of the unreliability of the subdiagnostic categories (Wehlman, 1952), the diagnoses of psychoses were not further broken down, although there appeared to be about the same proportion of affective psychotic states in both of the predrug and drug periods.

The conclusions regarding the over-all effectiveness of the ataractic drugs on the rehabilitation of the psychotic patient are based on both a comparison of the median length of hospitalization and the return rates of predrug patients and drug patients. The median length of hospitalization demonstrates the efficiency of the hospital system, but not necessarily the effectiveness of treatment. The return rates within a year following discharge of both the predrug and drug groups are used, therefore, to evaluate the effectiveness of ataractic drugs on the over-all rehabilitation of the psychotic patient. Since waiting for one year after discharge would have involved a delay of several years, the follow-up data were based on the 75% of patients who left the hospital within 292 days of admission in the predrug group, and the 75% of drug group patients who had left the hospital by 252 days from the time of admission. This 40-day difference between the time it took 75% of

the predrug and 75% of the drug group to be discharged represented a difference of only four patients. In other words, on the 252nd day from the date of admission, 150 1956-57 patients had been discharged, while 146 1953-54 patients had been discharged. Thus, the rates of return to this hospital are based on a total of 150 1953-54 patients and 150 1956-57 patients. The return rates to this hospital were found to underestimate the return rates of psychotic patients to this and other hospitals within one year of discharge by 7%.² A careful personal contact follow-up of almost 100 psychotic patients revealed that of the patients who were rehospitalized within one year after discharge, the vast majority returned to the discharging hospital.

The length of hospitalization and return rates to the hospital may be related to population characteristics and not necessarily to the type of treatment received. On all 400 patients in this study, data with regard to marital status, residence, service connection, type of admission, etc. were analyzed in relation to length of hospital stay and return to the hospital within one year of discharge. Married patients were found to leave the hospital significantly sooner than single patients. Out of state residents remained significantly longer than residents. An examination of the 1953-54 and 1956-57 psychotic patient groups revealed that 47% of the 1953-54 admissions were married and 32% of the 1956-57 patients were married. On the other hand, 2% more of the 1953-54 group were out of state residents. Consideration of both these factors indicated that the

² This follow-up is part of the research procedure of the VA Psychiatric Evaluation Project.

1953-54 group was likely to remain in the hospital a shorter period of time, based on the difference in the population characteristics alone. In the later presentation of the median length of hospitalization figures, 134 1953-54 patients are matched on marital status and residence with 134 1956-57 patients. This matching allows for a more accurate estimate of the effect of the introduction of ataractic drugs on the length of hospitalization. None of those variables were related to the return rates. The married patient, for instance, was just as likely to return to the hospital within a year of discharge as the single patient.

Results

From data presented in Table 1, there were no statistically significant differences in the median length of hospitalization of 200 pre-drug and 200 drug psychotic admissions, nor in the return rates of the 150 1953-54 (pre-drug) and 150 1956-57 (drug) discharged patients. At any one given time an average of 3% (total group) to 5% (matched group) more drug patients had been discharged from the hospital compared with the number of pre-drug patients. For example, within 100 days of admission, 86 1953-54 patients had been discharged, while 92 1956-57 patients had been discharged from the hospital. On the average of 6 to 7 days later, however, an equal number of 1953-54 patients had been discharged from the hospital.

The *direction* of the differences in Table 1 suggests that the ataractic drugs resulted in a more efficient hospital operation but a less effective treatment program.³ Since these differences were not significant, however, it was necessary to conclude that the use of ataractic drugs in the treatment of psychotic patients in a well staffed hospital has no demonstrable effect on the rehabilitation of the patient.

Discussion

The tangible within-hospital results of the use of ataractic drugs, however, are apparent. The wards are quieter, ECT has been reduced from 54 patients treated in 1953-54 to 16 patients treated in 1956-57. These changes have made it a more comfortable place for

the staff to work, and perhaps a more desirable place for patients to come. The fact remains, however, that the tendency of this well staffed hospital to rely so heavily on the ataractics has resulted in no measurable increase in over-all patient rehabilitation. On rounds and in staff conferences, the patient who is not improving is usually treated by a change in the dosage of his medication or by a change from one drug to another.

One possible explanation of the suggested higher return rate for patients treated with drugs is that the hospital was admitting more chronic patients in 1956-57, and, therefore, could expect a higher return rate following discharge. In a recent study on the rehabilitation of "hard core" chronic patients who had been hospitalized elsewhere at least five years prior to transfer (Ellsworth, Mead, & Clayton, 1958), it was reported that over 50% of these patients, with extremely poor prognoses, had been discharged from the hospital. The return rate, to *any* hospital within one year after discharge, was 31%. Since the return rate to *this* hospital reported for admission patients treated in the drug era (1956-57) was 31.3%, it would appear that chronic patients have at least as good a chance of remaining out of the hospital for one year as the more acute admission patients. It must be remembered that the 31.3% return rate of the 1956-57 psychotic to this hospital underestimated the actual return rate to this and other hospitals by 7%. In light of this evidence, the conclusion that there is no essential difference in patient rehabilitation as a result of the introduction of ataractic drugs is maintained.

Why are these drugs used and relied on so heavily in a well staffed psychiatric hospital? From the data of the present study, it is apparent that patients improved before the advent of ataractic drugs at about the same rate as patients treated with these drugs. The heavy utilization of these drugs reflects the staffs' conviction of their effectiveness, which can perhaps best be attributed to the error of *post hoc ergo propter hoc* logic. In other words, the staff assumes that since many patients improved following the administration of an ataractic drug, the drug must be the causative agent of that improvement. Actu-

³ Many of the 1956-57 patients were discharged and continued taking medication after leaving the hospital (based on trial visit reports). Even this additional factor of posthospital medication apparently made no difference in the over-all return rate of this group.

ally, of course, the administration of the drug was only one of several changes in the patient's life following his admission to the hospital.

Summary

1. The length of psychiatric hospitalization and return rates of two large groups of psychotic, male, nongeriatric patients admitted to a well staffed hospital were compared. One group (1953-54) was treated before the introduction of ataractic drugs. The other group, over 90% of whom were given ataractic drugs, was treated during the period 1956-57.

2. Aside from such in-hospital effects as quieter wards and less ECT, there were no statistically significant differences in the median lengths of hospitalization and the return rates of the two groups.

3. A careful examination of the differences in population characteristics of patients admitted during both years failed to account for the similarity of release and return rates during the drug and nondrug era.

REFERENCES

- ANDERSON, R. A., & FECHNER, A. H. Preliminary appraisal of hospital design. *Amer. J. Psychiat.*, 1956, 113, 211-217.
- ELLSWORTH, R. B., MEAD, BEVERLY T., & CLAYTON, W. H. The rehabilitation and disposition of chronically hospitalized schizophrenic patients. *Ment. Hyg. N. Y.*, 1958, 42, 343-348.
- STAUDT, VIRGINIA M., & ZUBIN, J. A biometric evaluation of the somatherapies in schizophrenia. *Psychol. Bull.*, 1957, 54, 171-196.
- WEHLMAN, B. The reliability of psychiatric diagnosis. *J. abnorm. soc. Psychol.*, 1952, 47, 577-578.

(Received January 29, 1959)

A COACTION COMPASS EVALUATION OF RORSCHACH DETERMINANTS IN BRAIN DAMAGED INDIVIDUALS¹

ANN LODGE GOTTLIEB AND OSCAR A. PARSONS

Duke University Medical Center

The purpose of this study is to predict the performance of brain damaged individuals on the Rorschach test by utilizing a recently developed theoretical approach to Rorschach analysis. This approach is based on the application of the Haskell Coaction Compass (1949) to personality dynamics and has proved fruitful in a number of studies (Clarke & Lodge, 1955; Lodge, 1953; Lodge & Gibson, 1952; Lodge & Guyer, 1953; Lodge & Steenbarger, 1953). The application of Haskell's Coaction method to psychology is based on the observation by Lodge and Steenbarger² that:

Personality theory, historically as well as recently, has inclined to rest largely upon assumptions of dynamic interplay between two sets of intrapsychic power systems. . . . Such standard dichotomies as thought and feeling, reason and instinct, ego and id, inhibition and facilitation, socialized and primitive reactions, "consciousness" and "unconsciousness," etc., all suggest interdependent power systems of roughly comparable meaning.

Interpretation of the Rorschach test, derived as it was from Rorschach's views on Jung's psychoanalytic formulations, has also traditionally relied on the assumption of two such intrapsychic power systems. Lodge and Steen-

barger (1953) have identified these variables as Control and Affect and provided operational definitions for them in terms of Rorschach responses.

The determinants of any given Rorschach response are considered to reflect the interaction of these two variables, which are defined so that form level represents the strength of the control component, while color, shading, vista, and movement determinants are considered expressive of affect.³ Three levels of each are specified: weak (-), neutral (0), and strong (+), thus yielding nine possible qualitatively different relationships or types of coaction.

Relationships between these variables can be visually represented conveniently on a cartesian coordinate system called the coaction compass (Fig. 1), whose *X* and *Y* axes represent, respectively, Control and Affect. Each response becomes a resultant vector of unit length, with its direction determined by the relative strength of the two coacting processes. All response vectors in a given protocol may be geometrically combined to yield a single overall vector representing the determinant pattern of the entire record. This vector may fall in any direction on the compass, due to the infinite number of possible intergradations between the nine coaction categories, and its length may range between 0 and 100% of the radius (the longer the vector, the more consistent is a given response tendency and vice versa).

³ The assumption of the orthogonality or independence of the two basic variables of control and affect has been supported by the finding of substantially zero correlations between them in a representative sample of protocols selected from previous coaction studies.

¹ An earlier version of this paper was presented at the Southeastern Psychological Association Meeting, Atlanta, Georgia, April 29, 1958. This research was supported, in part, by grant B-1459 from the National Institute of Neurological Diseases and Blindness, United States Public Health Service. The authors gratefully acknowledge the helpful suggestions of G. T. Lodge.

² From "Relations of the Personalities of Scientists, Normal Persons, Neurotics, and Psychotics: A Geometric Analysis," by G. T. Lodge and C. J. Steenbarger to be published in Volume II of E. F. Haskell (Ed.), *An Introduction to Unified Science*.

Vectors falling in the right half of the compass reflect good form level, those in the left half, poor form level. Vectors falling in the upper half of the compass reflect the presence of affect, those in the lower half result from the absence of affect determinants. Axes of Integration and Rigidity are defined by bisecting the right angles formed by the control (X) and affect (Y) axes. Positive integration ($++$) thus refers to strong control combined with strong affect, while weak control together with weak affect ($--$) defines its negative, i.e., lack of integration. Similarly, Rigidity ($+-$) refers to the ratio reflecting strong control with weak affect, while its opposite, Fluidity ($-+$), refers to weak control with strong affect.

Previous studies have demonstrated the effectiveness of coaction technique in differentiating the records of normal, superior (scientists), neurotic and psychotic individuals (Fig. 1) (Clarke & Lodge, 1955; Lodge, 1953; Lodge & Gibson, 1953; Lodge & Guyer, 1953; Lodge & Steenbarger, 1953). Briefly, some of the conclusions of these studies are (Clarke & Lodge, 1955; Lodge & Steenbarger, in press): (a) Psychologically healthy individuals, by their greater strength in *both* the control and affect components, can be differentiated from the psychologically ill. (b) Weakness in the affect component clearly differentiates neurotics as a group from both normal and superior individuals. (c) Weakness in the control component definitely distinguishes hebephrenics and catatonics from both normals and neurotics. Most psychotics also show a deficit in level of affect. (d) There is a continuum of psychological health corresponding to the integration axis scores according to coaction principles this effectively differentiates individuals in terms of the continuum.

The present study was initiated to further explore the value of this tool in personality research by testing its predictive value with respect to the performance of another group, i.e., brain damaged. In light of the previously cited results of earlier coaction research, known psychological effects of brain damage (Goldstein, 1942) and other Rorschach studies of the brain damaged (Fisher & Gonda, 1955; Klopfer & Kelley, 1946; Peña, 1953;

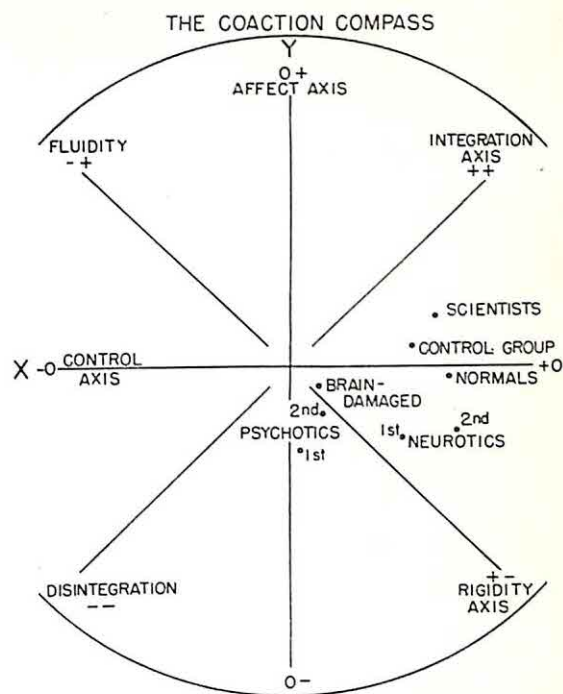


FIG. 1. Mean vector tip positions for brain damaged and control Ss, as well as original and cross-validation population samples of normals, neurotics, and psychotics (Clarke & Lodge, 1955).

Piotrowski, 1937, 1940; Reitan, 1955; Schreiber & White, 1954), two interrelated hypotheses were developed.

1. Because cortical damage often results in intellectual and perceptual impairment, a relative deficit in form level is predicted, with a consequent low score on the control or X axis for a brain damaged (BD) group.

This postulated deficit in form level or, at a behavioral level, control, is further assumed to result in a disturbed balance between forces of control and affect in the personality of the BD individual. Behaviorally, this disturbance might be expected to take the form of relatively consistent overcontrol, extreme lability, or fluctuations between rigid control and uncontrolled emotional outbursts. In the analogous Rorschach situation, such problems of behavioral integration would result in unsuccessful combination of control and affect determinants, i.e., minus form level when form is used in combination with other determinants (e.g., $CF-$, $FC-$), and the use of only one kind of determinant, such as F alone or pure C . These considerations lead to the second hypothesis.

TABLE 1
AGE, SEX, AND EDUCATIONAL CHARACTERISTICS OF CONTROL AND BD SUBJECTS

Subjects	N	Age		Sex		Education	
		\bar{X}	Range	M	F	\bar{X}	Range
Controls	20	34.0	15-55	10	10	10.9	3-18
Brain damaged	20	31.6	15-54	10	10	9.9	5-14

2. BD individuals will have a shorter projection on the axis of integration (X').

There would seem to be little psychological basis for a differential prediction concerning the expression of affect determinants in a BD group other than to anticipate considerable variability reflecting the reaction of different personalities to various kinds of damage.

METHOD

The basis of selection for both the BD and Control groups was referral to the psychologist for aid in differential diagnosis with the question of cortical damage as one possibility. As Piotrowski (1940) has pointed out, this type of control group would appear to be the most relevant. From a pool of 80 patients thus referred, 20 patients with unequivocal brain damage were matched for age, sex, color, and education with 20 patients whose ultimate medical diagnosis ruled out brain damage (Table 1). The criteria for inclusion in the BD group was either: (a) operational removal of a tumor or lesion, (b) the presence of one unequivocal neurological sign (such as known history of CVA, brain tumor, etc.), or (c) two or more neurologically equivocal signs where final medical diagnosis favored an organic cause. In accordance with the intended focus on possible general factors affecting BD behavior, there was a wide range of brain syndromes represented. Five of the 20 BD patients were diagnosed as having psychomotor and/or *grand mal* seizures alone. The remaining 15 manifested symptoms of acute or chronic brain syndrome following a variety of CNS disease and trauma, including, primarily, tumor removal, CVA, and cortical atrophy.

The criteria for the Control group consisted essentially of the absence of the three above conditions, although the presence of equivocal neurological signs was acceptable if final medical diagnosis was negative with respect to an organic cause. Also excluded were patients with diseases which are frequently associated with some ultimate cortical damage, such as hypertension, *diabetes mellitus*, or a significant degree of cardiac, pulmonary, or blood disease. The members of the Control group, thus determined, consisted largely of neurotic patients with considerable depression, anxiety, and psychosomatic or hysterical symptomatology. Five of the Control group patients were diagnosed schizophrenic. All information rele-

vant to the selection of subjects (S_s) was obtained from their medical charts rather than their psychological test records.

As seen in Table 1, the age range was 15 to 55 for both groups. The mean age of the BD group was 31.6 as compared with 34.0 for the Control group, a nonsignificant difference which, in any event, would operate against a prediction of greater deficit in the BD group. The difference between the mean levels of education, 9.9 for the BD group and slightly higher (10.9) for the Control group, was also insignificant.

All of these S_s had scorable Rorschach records in their psychological file. No restrictions were placed on the number of responses or rejections since this was considered to reflect realistically the clinical situation with respect to the types of cases under investigation. Although previous coaction studies have utilized Beck's scoring, Phillips Revised Developmental Scoring System (Phillips, Kaden, & Waldman, 1957) was used here. Phillips' system is considered particularly suited for the analysis of organic records because it gives credit for the positive elements in a percept whose over-all organization is inadequate. In adapting Phillips' definitions of adequacy and specificity of form to coaction levels of +, 0, and -, the only significant departure from previous studies was to equate responses scored "vague" in Phillips' system with 0 form level on the coaction compass. Since some previous studies have emphasized the prevalence of vague W responses in protocols of organics, it seemed this might help differentiate the records (Peña, 1953; Schreiber & White, 1954). It should also be noted that Phillips' system tends to give credit for a greater number of determinants as a result of the conditions under which a response is broken down into separately scored percepts. In scoring + and - form level, Phillips' revised rules were supplemented by the tables in Phillips and Smith (1953).

The records were rescored in the Phillips system by the senior author, using the original scoring by the examiners of these patients as a guide for determinants and form level ratings. The coaction compass vector for each record was then computed according to the method outlined by Lodge and Steenbarger (1952). To guard against possible bias due to knowledge of the composition of the groups as well as to establish the reliability of the scoring system, the 40 records of both the BD and Control groups were again scored by a research technician familiar with the Phillips system but unfamiliar with the in-

tent of the study and the nature of the population.⁴ The reliability scoring of the variables which affect coaction analysis revealed 92% agreement per protocol regarding the definition of responses, 99% agreement concerning the presence of determinants other than form, and 88% agreement on the scoring of form level. This degree of agreement was essentially the same for both groups and indicates satisfactory reliability of scoring.

As a further estimate of the representativeness of the Ss, the records were rated for Piotrowski's signs (1937). The BD group averaged 3.9 Piotrowski signs per record, with 9 records scoring beyond the recommended cutoff for diagnosing organicity of five or

⁴ The authors gratefully acknowledge the assistance of Shirley Crooke, who carried out the reliability study.

more signs. The Control group averaged 1.95 signs per record, with none showing more than four signs. This detection of 9 out of 20 "true" positives and no "false" positives is similar to Fisher and Gonda's findings (1955).

RESULTS

Figure 1 illustrates the resultant mean vector positions with respect to the *X* and *Y* axes for both the BD and Control groups. For the Control group, the mean projection on the *X* axis was 35.84 (percentage of radius). This is similar to projections previously reported for neurotic groups of 32.59 and 48.56 (see Table 2). However, this group showed consid-

TABLE 2

SUMMARY OF THE MEANS, SIGMAS, AND SIGMAS OF THE MEANS OF PROJECTIONS UPON THE AXES OF CONTROL AND AFFECT, INTEGRATION AND RIGIDITY, FOR BRAIN DAMAGED AND CONTROL GROUPS OF THE PRESENT STUDY AS WELL AS ORIGINAL AND CROSS-VALIDATION SAMPLES^a

(The units used in these tables are percentages of the radius of the coaction compass.)

		Control	Affect	Integration	Rigidity
Brain damaged (<i>N</i> = 20)	Mean	8.05	- 6.14	1.54	-10.00
	Sigma Dis.	21.28	36.14	27.94	31.43
	Sigma Mn.	4.88	8.29	6.41	7.21
Control group (<i>N</i> = 20)	Mean	35.84	6.65	29.14	-21.51
	Sigma Dis.	4.12	20.28	18.92	15.89
	Sigma Mn.	.94	4.65	4.34	3.64
Original data Psychotics (<i>N</i> = 25)	Mean	2.54	-25.21	-16.10	19.66
	Sigma Dis.	26.03	28.54	24.27	30.05
	Sigma Mn.	5.31	5.82	4.95	6.13
Neurotics (<i>N</i> = 27)	Mean	32.59	-20.81	8.33	36.61
	Sigma Dis.	10.68	16.07	12.25	15.69
	Sigma Mn.	2.10	3.15	2.40	3.08
Scientists (<i>N</i> = 19)	Mean	42.65	15.84	41.35	18.95
	Sigma Dis.	5.73	20.03	14.98	14.50
	Sigma Mn.	1.39	4.86	3.64	3.52
Cross-validation data Psychotics (<i>N</i> = 30)	Mean	9.22	-13.88	- 3.29	16.33
	Sigma Dis.	15.17	26.43	22.03	21.05
	Sigma Mn.	2.77	4.83	4.02	3.84
Neurotics (<i>N</i> = 30)	Mean	48.56	-18.95	20.93	47.75
	Sigma Dis.	12.63	22.82	17.41	19.44
	Sigma Mn.	2.31	4.17	3.18	3.55
Normals (<i>N</i> = 30)	Mean	46.39	- 2.83	30.80	34.82
	Sigma Dis.	13.16	19.86	18.51	15.01
	Sigma Mn.	2.40	3.63	3.38	2.74

^a Clarke and Lodge, 1955.

erably less variation than the previous groups with a standard deviation of only 4.12. This suggests that as regards form level and its suggested analogue of control, these Control Ss (who were chosen on the basis of a certain behavioral similarity, namely, organiclike symptoms) are a relatively homogeneous population in terms of this personality factor.

The mean projection on the X axis for the BD group was 8.05 with a sigma of 21.28. Twelve BD Ss scored below the lowest Control S's score of 12.5, and only one BD S scored above the Control group mean. The Mann-Whitney U test was utilized to test the hypothesis that the two groups represented different populations with respect to this dimension. The results supported the initial prediction of a lower score on the Control axis for the BD group with $p < .001$. In comparison with results of earlier studies, the BD are most similar to psychotics (functional) on this dimension (Table 2).

The second prediction, that the BD group would show a shorter projection on the Integration axis, was also supported with a $p < .001$ according to the results of the Mann-Whitney U test. Table 2 indicates a mean of 1.54 for the BD and 29.14 for the Controls in this axis. The distribution was such that only five Control Ss scored below the median projection of 19.5 on the axis of integration, while only five BD Ss scored above this median.

With regard to a breakdown of the response types that contributed to these two results, it is interesting to note that there were no significant differences between the groups in the number of vague responses. The prevalence of minus form level in the BD records, both in responses determined by form alone and combined with other determinants, contributed most to the present findings. It should be noted, however, that the final projection on the Control axis represents a different kind of information than that given by the $F+$ conventionally used in Rorschach scoring, since every response is weighted either +, -, or 0, for both form and affect. Of the BD records, 70% contained at least one W minus response, whereas this sign was present in only 20% of Control group records, a difference significant beyond the .01 level. Incidentally, the total number of responses given

by both BD and Control Ss was unexpectedly similar, with both groups averaging 14.0 responses per protocol.

With respect to affect determinants, the only difference between the groups was the anticipated greater variability among the BD scores, which was significant at the .02 level of confidence according to an F test applied to the difference between the variances. Previously tested psychotic groups also showed wide variability on the affect dimension. The mean projection on the Y axis was +6.65 for the Control group and -6.14 for the BD. The tendency of the Phillips system to recognize a greater number of determinants may contribute to the higher score of this Control group in comparison with previous norms for neurotic groups (Table 2).

DISCUSSION

The finding of significantly lower scores on the Control or X axis for the BD group confirms the first hypothesis and suggests that cortical damage is associated with relatively inadequate form perception. This kind of impairment may be interpreted as symptomatic of a more general deficit in ego strength (Phillips & Smith, 1953) or in coaction terms, control, and related to inadequacies in reality testing.

The second hypothesis concerned the Integration of what has been defined as control with factors considered to reflect emotional responsivity. Its confirmation by a significantly shorter projection on the Integration axis for the BD group is largely due to the poor form level. This leads to the generalization that weakened control at a behavioral level is the primary factor leading to a disturbed balance of control and affect components of personality and their resultant inadequate integration. The wide range of scores along the affect dimension for the BD group suggests that this leads to a variety of emotional adaptations. On the whole, the data obtained indicate that the BD group is more similar to psychotic groups than to either normal or neurotic populations. This similarity is interesting in light of the fact that patients with functional psychoses frequently manifest organiclike symptoms.

In view of the fact that coaction studies

thus far have focused largely on comparing the Rorschach performance of different diagnostic groups, further implications of these findings would require research correlated with more specific behavioral assessment. For example, a promising area for future research suggested by the results of the present study might involve investigation of the correspondence between scores on the integration-disintegration and fluidity-rigidity continua with actual behavioral manifestations of inadequate integration of control and affect in brain damaged individuals.

At present, coaction methodology represents primarily a systematization of Rorschach determinant scoring. This leads to the general question of why, in the absence of clearly specified behavioral correlates related to control and affect, have coaction studies consistently yielded results which significantly differentiate among healthy groups as well as among various categories of psychopathology. One hypothesis is that coaction analysis represents a geometric encoding of certain implicit inferential processes involved in the diagnostic behavior of clinicians. Lodge and Steenbarger (see fn. 2) have implied this in their suggestion of the fundamental equivalence of major personality theory concepts which tend to rest on the basic assumption of a two-variable coacting intrapsychic power system. It may be that Kraepelinian and psychoanalytic models, despite differences in terminology, tend implicitly to utilize a common logic to analyze, differentially weight, and finally integrate information into diagnostic statements. Disregarding, for the moment, the problem of the validity of commonly used categories (for example, the distinction between neurotic and psychotic), there is the possibility that the formal geometric system imposed upon Rorschach data by coaction scoring represents a significant approximation to the principal common elements in the logic of diagnosis.

If there is, then, some theoretical justification for exploring possible interactions between two such variables, for which control and affect are convenient approximations, the general framework of a cartesian coordinate system permits representation of the possible interactions between them according to a con-

sistent logical system. As a geometric model, the *two-dimensional* coordinate system has the advantage over traditional verbal conceptual frameworks of avoiding the possible imposition of unidimensionality upon the interactions of a *two-variable* system. It also conveniently permits visual comprehension of these numerous *relationships*, *simultaneously*. At the same time, of course, it involves a network of geometric conventions. These, however, lead to explicit hypotheses which are subject to test. The relation of this geometry to the problem of diagnosis, its language system and empirical referents, then, are areas of research suggested by the results of this and other coaction studies. In view of the present lack of unification in the language system of psychopathology, especially in its relationship to concepts of mental health, it would seem crucial to clarify whatever underlying logical framework may exist.

SUMMARY

Coaction analysis of Rorschach divides determinants into two variables: Control (form level) and Affect (color, shading, and movement). Interactions between these determinants are geometrically plotted to yield a resultant vector on a cartesian coordinate system. On the basis of symptoms described in the brain damaged, it is predicted that they will manifest a breakdown of form level in combination with other determinants to result in (a) a lower score on the Control axis and (b) a shorter projection on the Integration axis.

From a pool of 80 patients referred for psychological testing with questionable organic brain disease, 20 patients with unequivocal brain damage were matched on age, sex, color, and education with 20 patients whose ultimate medical diagnosis ruled out brain damage. The Rorschach records of these patients were scored according to the Phillips Revised Rorschach Developmental Scoring System and then plotted on the coaction compass.

The results confirm the predictions and indicate that the BD population is similar to previously tested psychotic groups both in terms of weakened control, poor integration

of control and affect, and considerable variability along the affect dimension.

REFERENCES

- CLARKE, J. I. C., & LODGE, G. T. Coaction compass positions of normal, neurotic, and psychotic individuals: A cross-validation. *Amer. Psychologist*, 1955, 10, 325. (Abstract)
- FISHER, J., & GONDA, T. A. Neurologic techniques and Rorschach test in detecting brain pathology: A study of comparative validities. *AMA Arch. Neurol. Psychiat.*, 1955, 74, 117-124.
- GOLDSTEIN, K. *Aftereffects of brain injuries in war*. New York: Grune & Stratton, 1942.
- HASKELL, E. F. A clarification of social science. *Main Currents modern Thought*, 1949, 7, 45-51.
- KLOPPER, B., & KELLEY, D. M. *The Rorschach technique*. (Rev. ed.) Yonkers, N. Y.: World Book, 1946.
- LODGE, G. T. A method for the dynamic representation of personality data. *Amer. Psychologist*, 1952, 7, 348 (Abstract); *J. proj. Tech.*, 1953, 17, 477-481.
- LODGE, G. T., & GIBSON, R. L. A coaction map of the personalities described by H. Rorschach and S. J. Beck. *J. proj. Tech.*, 1953, 17, 482-488.
- LODGE, G. T., & GUYER, E. G. Scientists and sanity. *Amer. Psychologist*, 1953, 8, 390 (Abstract); *Gen. Semantics Bull.*, 1955 (Nos. 16-17), 35-39.
- LODGE, G. T., & STEENBARGER, C. J. Forms for the computation of the Rorschach coaction vector. 1952. (Mimeo.)
- LODGE, G. T., & STEENBARGER, C. J. Charting the course of the Rorschach interview. *J. gen. Psychol.*, 1953, 48, 67-73.
- PEÑA, C. D. A genetic evaluation of perceptual structuring in cerebral pathology: An investigation by means of the Rorschach test. *J. proj. Tech.*, 1953, 17, 186-199.
- PHILLIPS, L., KADEN, S., & WALDMAN, M. The application of developmental theory to problems of social adaptation. 1957. (Dittoed manual)
- PHILLIPS, L., & SMITH, J. G. *Rorschach interpretation: Advanced technique*. New York: Grune & Stratton, 1953.
- PIOTROWSKI, Z. A. The Rorschach Ink-Blot method in organic disturbance of the central nervous system. *J. nerv. ment. Dis.*, 1937, 86, 525-537.
- PIOTROWSKI, Z. A. Positive and negative Rorschach organic reactions. *Rorschach res. Exch.*, 1940, 4, 147.
- REITAN, R. M. Validity of Rorschach test as measure of the psychological effects of brain damage. *AMA Arch. Neurol. Psychiat.*, 1955, 73, 445-451.
- SCHREIBER, HANNA, & WHITE, MARY ALICE. Diagnosing organicity on the Rorschach. *Psychiat. Quart. Suppl.*, 1954, 28, 255-277.

(Received February 9, 1959)

INTRA-Q DECK RELATIONSHIPS AS INFLUENCES AND REALITIES IN PERSONALITY ASSESSMENT¹

PETER F. BRIGGS AND ROBERT D. WIRT

Medical School and Center for Personality Research, University of Minnesota

Statistical controls are thought to provide a means of convergence between the ambiguous and the known. In the analysis of data drawn from Q sorts it has been found that stereotypes such as sick-well (Kogan, Quinn, Ax, & Ripley, 1956) and social desirability-undesirability (Edwards, 1957) can be identified. The suggestion is then made that the effects of these stereotypes be controlled either by item selection or by correlational techniques. These recommendations rest upon work aimed at identifying stereotypes that parallel a good-bad dimension. Much has been made of the ease with which personality descriptive items can be ordered on such dimensions. The actual validity of such descriptions has not been examined extensively.

The Q sort technique produces an amalgamation of sorter, subject, and Q deck characteristics which are difficult to refract. The group or case descriptions resulting from these multiple influences are usually considered primary data; that is, they are rarely checked against other descriptions of the same cases.

This report includes an examination of Q deck stereotype descriptions as they relate to (a) known characteristics of samples, (b) each other, (c) item variance, and (d) reliable group differences as reflected by the items in the deck.

PROCEDURE

The cases. Four groups were examined. Each had approximately 70 cases who were selected because of their Minnesota Multiphasic Personality Inventory (MMPI) patterns and their delinquency status. Two

¹ This study was supported by research grants from the graduate school of the University of Minnesota and the National Institute of Mental Health, National Institutes of Health, United States Public Health Service.

samples had MMPI codes that contained a peak on *D*, or *Mf*, or *Si*, the scales which other research has shown to be associated with a low rate of delinquency (Hathaway & Monachesi, 1953; Wirt & Briggs, 1959). One of these groups had no record of delinquency, and one was composed of delinquents. Another pair of delinquent and nondelinquent samples had MMPI high point codes containing a two-scale combination of *Pd*, or *Sc*, or *Ma*. These are scale combinations which have been shown to be associated with delinquency proneness (Hathaway & Monachesi, 1953; Wirt & Briggs, 1959). The two obvious characteristics of the groups were the delinquency versus nondelinquency dichotomy, and the personality test differences, which suggest a neurotic versus character disorder dichotomy. The groups also differed greatly in achievement (school and job success) as well as other forms of constructive community activity. The neurotic nondelinquent sample was the most successful group on these criteria of success. The two delinquent samples were indistinguishable in achievement and clearly below both nondelinquent samples.

The rating of each sample as well as the collection of additional information has been reported elsewhere (Wirt & Briggs, 1959). Cases from each sample were seen for intensive interviews. On the basis of these, the interviewers sorted a Q deck for each S.

The Q deck. The deck used in this study was developed by Block (1954). It contained 115 items to be distributed over a forced nine-point distribution. Two aspects of item variability were examined. These were the size of the variance when an item was used as descriptive of a group, and the relation of items showing a significant difference among group means.

The interviewers. The interviewers were five graduate students in psychology or social work, all of whom had had at least three years of clinical experience.²

Stereotype sorts. Following the final interview and Q sort, the interviewers were instructed as follows:

Using your experience with the deck for this sample of cases, sort the Q deck for the five concepts listed below: (a) social desirability of the traits in

² The interviewers in this investigation were Douglas Freed, Arthur Gallese, Ronald Johnson, A. M. Marchionne, and Jay Willet.

TABLE 1
MEAN INTERCORRELATIONS FOR EACH STEREOTYPE
AMONG FIVE SORTERS

Stereotype	N	Mean r
Delinquent	3	.32**
Nondelinquent	3	.73**
Neurotic	3	.63**
Social desirability	5	.76**
Item difficulty	5	.41**

** $P < .01$.

the deck; (b) difficulty of sorting each item in the deck; (c) impression you formed of the typical delinquent; (d) impression you formed of the typical nondelinquent; (e) impression you formed of the typical neurotic.

RESULTS

The mean Pearsonian correlation coefficients among sorters for each of the five sorts were

calculated using Fisher's z transformation. They are given in Table 1. The concept of delinquency was poorly agreed upon ($r = .32$). For this reason it was dropped from subsequent analyses even though there was some reliable covariance. The agreement on difficulty in sorting was not encouragingly high. The remaining values appeared to be respectable interjudge correlations.

Mean item ratings for the four samples and the stereotypes were intercorrelated. These results are given in Table 2. The matrix of intercorrelations in Table 2 has three subdivisions: the interstereotype correlations, the intersample correlations, and correlations between the stereotype sorts versus the mean sample sorts.

The four samples were compared with each other in their relationship to the three stereo-

TABLE 2
CORRELATION COEFFICIENTS BETWEEN SAMPLE AND STEREOTYPE ITEM MEANS

Sample or Stereotype	Neurotic		Character Disordered		Stereotype			
	Non- delin- quent	Delin- quent	Non- delin- quent	Delin- quent	Non- delin- quent	Neurotic	Social Desira- bility	Item Diffi- culty
Neurotic nondelinquent sample	—	.72**	.72**	.41**	.67**	.21**	.52**	.06
Neurotic delinquent sample	—	—	.82**	.83**	.39**	-.08	.61**	.09
Character disordered nondelinquent sample	—	—	—	.72**	.84**	-.28**	.77**	.18
Character disordered delinquent sample	—	—	—	—	.61**	-.32**	.55**	.11
Nondelinquent stereotype	—	—	—	—	—	-.34**	.86**	.25**
Neurotic stereotype	—	—	—	—	—	—	-.43**	-.20**
Social desirability stereotype	—	—	—	—	—	—	—	.33**

* $P < .05$.

** $P < .01$.

TABLE 3
SIGNIFICANT SIMILARITIES BETWEEN STEREOTYPES AND SAMPLES

Samples Compared	Stereotype		
	Nondelinquent	Neurotic	Social Desirability
	Like Sample	Like Sample	Like Sample
Neurotic nondelinquent (I) vs. neurotic delinquent (II)	I**	II**	—
Neurotic nondelinquent (I) vs. character-disordered nondelinquent (III)	III**	I**	III**
Neurotic delinquent (I) vs. character disordered delinquent (IV)	—	I**	—
Neurotic delinquent (II) vs. character-disordered nondelinquent (III)	III**	II*	III**
Neurotic delinquent (II) vs. character-disordered delinquent (IV)	IV**	II**	—
Character-disordered nondelinquent (III) vs. character-disordered delinquent (IV)	III**	—	III**

* $P < .01$.

** $P < .001$.

types having the highest interrater agreement: nondelinquency, neuroticism, and social desirability. These results are shown in Table 3. The validity of these stereotypes was then assessed in terms of known characteristics of the samples.

Nondelinquent stereotype and known nondelinquency. The samples of nondelinquents were accurately described as less nondelinquent than the delinquent samples with similar personality test patterns. The nondelinquent neurotic sample appeared about the same as the character disordered delinquent sample. In this case the stereotype did not distinguish the known group behaviors. However, on the other cross-personality pattern comparison a significant difference was found, showing the nondelinquent character disordered group to be similar to the nondelinquent stereotype and the delinquent neurotic code group correctly was not described as nondelinquent. The surprising thing in these findings is the degree to which the character-disordered samples appeared more nondelinquent than the neurotic samples. Other data show that, in fact, the character-disordered samples consisted of individuals with a lower re-

gard for the mores than most individuals with test findings indicating neurotic disturbance. The nondelinquent stereotype apparently was complicated by other influences.

Neurotic stereotype and personality test findings. The results for the neurotic stereotype group comparisons also appear in Table 3. Here the neurotic and the character disordered groups were distinguished on all occasions. It is interesting to note that the delinquent neurotic sample was described as more neurotic than the nondelinquent sample with similar MMPI profiles. This suggests a surplus meaning in the stereotype. It should be noted that, in fact, the delinquent samples were actually more psychopathic on the MMPI and in actual behavior than their nondelinquent test pairs (Wirt & Briggs, 1959).

Social desirability stereotype and other criteria of social desirability. As indicated previously on the criteria of social desirability, school and job achievement, and constructive community enterprise, the neurotic nondelinquent sample was superior to all other samples. According to the stereotype developed by the raters, however, it was the character-

TABLE 4

THE NUMBER OF ITEMS JUDGED CHARACTERISTIC OF THE
SAMPLES AND STEREOTYPES WHICH OVER-
LAP THE EXTREMES OF SOCIAL
DESIRABILITY

Sample or Stereotype	Items			
	Socially Desirable (<i>N</i> = 19)		Socially Undesirable (<i>N</i> = 20)	
	Like	Unlike	Like	Unlike
Neurotic nondelinquent sample	6	0	0	11
Neurotic delinquent sample	9	0	0	6
Character disordered nondelinquent sample	10	0	0	11
Character disordered delinquent sample	9	0	0	7
Nondelinquent stereotype	12	0	0	13
Neurotic stereotype	0	5	4	4
Item difficulty stereotype	0	4	3	4

disordered nondelinquent sample that seemed most desirable. Further, known delinquency versus nondelinquency was not reflected in social desirability as sorted by the raters, although certainly delinquency is less socially desirable than nondelinquency. The relationship between the nondelinquent and social desirability stereotypes was the highest in the matrix ($r = .86$).

The MMPI and delinquency status were effective variables, as were the nondelinquent and the social desirability stereotypes. All of these correlated with the other samples and stereotypes in ways which tend to reflect the known group characteristics.

The high correlations associated with social desirability as an a priori stereotype accounts

for a quite uniform amount of variance in all samples except the sample of nondelinquent character disorder individuals. This sample approached the stereotype of "good" personality to a striking degree. It is important to note that this finding is not paralleled in other, less subjective, data where it was found that these individuals would not be considered socially desirable by conventional standards (they are unachieving, promiscuous, unreliable, etc.).

Removal of the variance attributable to social desirability would not appreciably alter the relationships in the sample matrix. An illustration of the overlap among all of the distributions underlying the matrix and the social desirability stereotype is provided in Table 4. Here the top and bottom 20 items in each sort were examined for overlap with the top 19 and bottom 20 of the social desirability items. It will be noted that in the nondelinquent stereotype and in the clinical groups there were no misplaced items. The correlations shown in Table 2 are, of course, a function of the frequency of these overlapping items. However, the concreteness of the examples in Table 4 provides a convincing illustration of the strength of the correlation.

The content of items which constituted the more desirable traits were intellectuality, emotionality, independence, and perspective. The undesirable traits seemed much more complicated and contained longer, more involved descriptions. The traits rated undesirable were highly interpretive and referred to intrapsychic phenomena; e.g., hostility, absence of warmth, and poor ego development. The ends of the distribution for the nondelinquent stereotype sorts were much the same as those for the social desirability items.

The items from the extremes of the neurotic stereotype described insecurity and preoccupation with neurotic problems at the characteristic end, while at the nonneurotic extreme the items reflected an absence of concern with such thoughts and a happy aggressive mastery of situations. Again, the content of the negative (more neurotic) items was more complicated and the descriptions were longer than the items sorted as more desirable.

Content of items in relation to item difficulty. An examination of the traits that were

regarded as difficult or easy to sort showed the more difficult items seemed to refer to quite obscure facts, usually relating to covert data or complex interpersonal relations. The items rated as easily sorted, by contrast, referred to more observable ways of acting or simply to expressed verbal opinions. The frequencies of overlap between the extremes of this distribution and the social desirability dimension extremes were presented in Table 4. The correlation between these two mean sorts was low ($r = .33$) but significant ($P < .01$). The small number of overlapping items shown in Table 4 reflects the weak relationship.

Item variance and mean differences. A final pair of measures of item performance studied were the item variances within the samples and the occurrence of significant mean differences among the sample means. From the 115 items, 24 with low variance were selected (items with two sample variances out of four below 3.00). Seventeen high variance items were selected (items with two variances greater than 5.50). The proportion of discriminating items in these two sets was almost identical. There were six items making 10 discriminations in the low set and five items making 9 discriminations in the large variance set. The items for the extremes of social desirability were not evenly represented among the items at the extremes of variability. There were four such items in the 24-item set with low item variability and 10 in the 17-item set with high variability ($P < .01$). Of these latter 10 items, five were from each end of the social desirability continuum. It was clear that the items with large variances were more often from the extreme of social desirability than were items with small variance. As can be seen in Table 4, the two extremes of the social desirability stereotype behaved consistently (although the items were different).

The ease-difficulty of sorting was examined for the items in the two extremes of variation. There were 6 items of high variance and 11 with low variance. The easy to sort items were all high variance items (5 high and 0 low), while the hard to sort items were almost all low variance items (1 high to 11 low; $P < .01$). Larger variances were associated with easier sorting (and with social desir-

ability). Items in the extremes of sorting ease were, however, almost equally discriminating in group comparisons. The number of items and discriminations made by the items in each group mean comparison for the extreme items showed no differences. Neither the nondelinquent nor the neurotic stereotype extreme items behaved differently with respect to variance size. However, the high variance items were more frequent in the nondelinquent stereotype extremes than were the low variance items. This would be expected from the relationship between the nondelinquent and social desirability stereotypes. Neither set of extreme items was differently involved in group differences.

DISCUSSION

These data support the argument that stereotypes can be formed from Q deck items. This evidence does not, however, substantiate the idea that the stereotypes derived from a priori sorts are especially meaningful. This is particularly true of the social desirability concept. In the forced sort the content available in the items as subtopics determines the form a stereotype may take.

With this particular set of items a type of glib, almost psychopathic caricature emerges as the stereotype of the socially desirable personality. It is quite conceivable that with other item sets quite a different stereotype might have emerged. To some degree, the items in this stereotype showed a high overlap with all the sample descriptions. It is clear that the desirability continuum describes the nature of any group to the extent that no items from the undesirable extreme are reversed in group description. Thus the extremes of such a stereotype are apparently forced by the sorters' knowledge of people. This is certainly not the typical meaning of a stereotype in psychological research. It more closely parallels the meaning of a cliché. This is not to imply that the items at the extremes of the stereotypes do not discriminate between cases, for they manifest more than average variability in the present study.

Among the items in the deck some connection between sorting ease and the stereotype extremes appeared clearly. Here it is suggested that stereotyping hinges upon base-

belief relationships. The easily sorted items were useful in stereotyping while more obscure factors simply did not enter. Similarly the easier to sort items were more variable.

As to the intradeck relationships and the items describing groups, there were no obvious limitations or correlations. The distinguishing characteristics of the groups seemed to emerge from all types of items.

A final point is in order. Edwards (1957) and others have proposed that items reflecting stereotypes be removed or controlled. This is a difficult procedure, since to identify the particular influences to be removed requires a very extensive analysis of the deck. But more importantly, it appears that valuable items will certainly be deleted under such a system of selection.

SUMMARY

Q sort descriptions of several samples and stereotypes were made by five judges following intensive interviews. These descriptions were compared with known characteristics of the Ss. The mean correlation among sorters for each of the five sorts of delinquency, nondelinquency, neuroticism, social desirability, and item difficulty was significant. The sorters accurately described the nondelinquents and the neurotic samples. They described the delinquents as more neurotic than the nondelinquents and the character disorder samples as more nondelinquent than the neurotic samples. In fact other data showed the character disorder samples to be more psychopathic on the MMPI and in actual behavior than other groups. The raters described the character-disordered nondelinquents as the most socially desirable, while in fact this group was known

to have a generally poor regard for social conventions. Known delinquency versus nondelinquency was not reflected in social desirability as sorted by the raters.

Items which were difficult to sort included content requiring interpretation of covert behavior, while items rated as easily sorted referred to more observable ways of acting. Items with large variance within samples were more often from the extremes of social desirability than were items with small variance.

The data were interpreted to show that while the Q sort technique may be useful in developing definitions for stereotypes, the resulting concepts must be interpreted with caution. This seemed especially true of the social desirability stereotype. It was suggested that the removal or statistical control of items reflecting stereotypes requires extensive analysis and can result in deletion of valuable information.

REFERENCES

- BLOCK, J. *The IPAR Clinical Q-Set (Form II)*, 115 personality descriptive items. Berkeley: Inst. of Pers. Assessment & Res., 1954.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- HATHAWAY, S. R., & MONACHESI, E. D. *Analyzing and predicting juvenile delinquency with the MMPI*. Minneapolis: Univer. Minnesota, 1953.
- KOGAN, W. S., QUINN, R. C., AX, A. F., & RIPLEY, H. S. Some methodological problems in the quantification of clinical assessment of Q array. *J. consult. Psychol.*, 1956, 21, 57-62.
- WIRT, R. D., & BRIGGS, P. F. Personality and environmental factors in the development of delinquency. *Psychol. Monogr.*, 1959, 13(15, Whole No. 485).

(Received February 9, 1959)

THE LONG TERM STABILITY OF SELECTED RORSCHACH RESPONSES¹

JEROME KAGAN

Fels Research Institute, Yellow Springs, Ohio

Of the various instruments employed to elicit fantasy responses, the Rorschach has been both the most frequently used technique and the one subjected to the most empirical research. However, despite many years of research effort, a large body of social scientists still question Rorschach's basic hypotheses and the reliability and validity of Rorschach derived variables. The purpose of this paper is to demonstrate that selected Rorschach variables show low, but statistically significant, degrees of intra-individual stability over periods of three and six years during adolescence and early adulthood. Since important personality changes often occur during these years one would not expect extremely high reliability coefficients over such a long period of time. Thus, absence of stability for a specific variable would not mean that the variable might not be reliable for shorter test-retest intervals. However, demonstration of a statistically significant degree of stability, albeit low to moderate, would suggest that the variable was reflecting an enduring aspect of human functioning and deserved more intensive study.

There is some indication that determinant variables show low to moderate degrees of test-retest reliability when the interval between tests is less than a week (Kaplan & Berger, 1956; Epstein, Nelson, & Tanofsky, 1957). In the Epstein et al. (1957) research 16 college students were presented with a different series of specially designed ink blots twice weekly for five weeks. Kaplan and Berger administered the standard Rorschach stimuli four times to 28 subjects with an in-

terval of four days between each administration. In both studies movement responses showed a moderate degree of reliability with correlations in the forties and fifties. However, except for a longitudinal study of the consistency with which the 10 standard blots elicit specific responses (Ranzoni, Grant, & Ives, 1950), there is no evidence on the long term intra-individual stability of content or determinant variables.

Historically, initial interest in Rorschach responses was centered on the formal, perceptual dimensions of location and determinant. However, during the last decade there has been a rapidly growing interest in the motivational content of Rorschach responses, with particular emphasis on aggressive content (DeVos, 1952; Elizur, 1949; Finney, 1955; Kagan, Sontag, Baker, & Nelson, 1958; Rader, 1957; Schafer, 1954; Towbin, 1955; Wolf, 1957; Zubin, Eron, & Sultan, 1956). The present study, therefore, was particularly interested in the stability of aggressive content. The intense interest in aggressive content seems due, in part, to the fact that Rorschach stimuli have a strong tendency to elicit content which appears aggressive in meaning, e.g., dead bug, animals fighting, people arguing, knives, guns, bombs exploding, etc. Undisguised evidence of other dynamic motivational percepts (sex or dependency) appears with less frequency. That is, perceptions of objects or actions symbolic of sexual or dependent motives (genitals, breasts, dogs nursing, people kissing) are less frequent than imagery normally regarded as symbolic of aggression. Second, clinicians often must make important predictions about a patient's aggressive motives and the probability of occurrence of overt aggressive behavior. This practical consideration would direct the

¹ This research was supported, in part, by a research grant (M-1260) from the National Institute of Mental Health, United States Public Health Service.

clinician's attention to imagery which appeared to be symbolic of aggressive thoughts.

In addition to aggressive content, the stability of seven additional variables was explored. The complete list of variables included: (a) number of responses (*R*), (b) dynamic aggression (*Dyn. Agg.*), (c) static aggression (*Stat. Agg.*), (d) anatomy (*Anat.*), (e) dependent-oral imagery (*Dep.-Oral*), (f) human and human detail responses (*H*), (g) human movement (*M*), (h) animal movement (*FM*), and (i) sum movement (*sum M, FM, m*).

The stability of movement responses was investigated because aggressive content often involves a movement dimension and because many psychologists regard the movement response as one of the most important Rorschach variables (Klopfer, Ainsworth, Klopfer, & Holt, 1954; Phillips & Smith, 1953; Piotrowski, 1958). Human percepts (*H*) were analyzed in order to determine the differential stability of human percepts with movement versus human percepts with or without the presence of movement. Anatomy was scored because of its presumed diagnostic value for pathological conditions. Finally, dependent-oral imagery was scored in order to compare the stability of two motive categories (aggression and dependency) that play an important role in the functioning of both normal and patient populations. Although we would have liked to score for location, form quality, color, shading, and vista, the nature of the available data made this analysis impossible. The protocols were gathered over a long period of time by different examiners with varied inquiry procedures and it was impossible to score for location and most determinants.

METHOD

Subjects and Procedure

The subjects (*Ss*) were drawn from the Fels research population, which is a predominantly middle-class sample residing in southwestern Ohio. The fathers of these *Ss* were equally distributed among the following vocational classes: agriculture, skilled labor, tradesmen, and professional. The IQ scores of the *Ss* were all average or above with a mean Stanford-Binet IQ of 120 during the childhood years, and a mean Wechsler-Bellevue IQ of 123 during early adult life. For this study, 37 male and 38 female *Ss* were selected for whom two or three standard Rorschach protocols were available between 10

and 20 years of age. In addition, 52 of these *Ss* (30 males and 22 females) were adult members of a current assessment program which has been in progress for several years. The ages of the 52 *Ss* at the time of the adult assessment ranged from 19 to 29 with a median age of 25 years.

The longitudinal Rorschach protocols were based on the 10 standard Rorschach plates which were administered at medium ages of 10½, 13½, and 16½, with 30 of the males and 33 of the females having all three administrations. As mentioned above, the examiners differed over the years, and the comparability with respect to the inquiry could not be guaranteed. For this reason, only the initial, spontaneous verbalization of the *S* (the free association) was scored. In the adult assessment program a modification of the Rorschach stimuli was used which will be called the modified Rorschach. In this modified series, parts of the standard stimuli were covered with a template of white cardboard. Thus the *S* saw only a part of the original stimulus and gave only one response to each card. The 32 stimuli were common *D* or *d* areas, each tended to be perceived as a unitary Gestalt, and, with one exception, all the stimuli were either completely achromatic or chromatic. This special instrument was designed for two reasons. First, it guaranteed an equal and fairly large response pool for all *Ss*. The statistical problems associated with unequal protocol length have long plagued Rorschach investigators, and this instrument was a beginning attempt to solve this thorny problem. Second, this instrument gives the investigator considerable control over the stimulus input to the *S* and allows the investigator to specify those aspects of the stimulus upon which the response was based. It is acknowledged that this alteration of the Rorschach stimuli produces a qualitatively different stimulus pattern to the *S*. However, the author was interested in ascertaining if specific types of response would show intra-individual stability despite alteration in the external stimulus. The 32 stimuli in this series were congruent with the following Klopfer areas (Klopfer et al., 1954, pp. 70-79).²

Rorschach Plate	Klopfer Areas
I	(<i>D</i> 1) (<i>D</i> 2) (<i>d</i> 3 symmetrical)
II	(<i>D</i> 1) (<i>D</i> 2 symmetrical) (<i>D</i> 3 symmetrical) (<i>d</i> 1)
III	(<i>D</i> 2) (<i>D</i> 1) (<i>D</i> 3) (<i>D</i> 8 symmetrical plus <i>D</i> 3)
IV	(<i>D</i> 1) (<i>d</i> 2) (whole blot minus <i>D</i> 1)
V	(<i>d</i> 2) (whole blot minus <i>d</i> 3 and <i>d</i> 1)
VI	(<i>d</i> 2) (<i>D</i> 4) (<i>d</i> 4)
VII	(<i>D</i> 4 symmetrical) (<i>d</i> 1)
VIII	(<i>D</i> 2) (<i>D</i> 1 plus <i>D</i> 3 plus <i>D</i> 4) (<i>D</i> 3 the center) (<i>D</i> 4 the bottom)
IX	(<i>D</i> 2 symmetrical) (<i>D</i> 5) (<i>D</i> 7)
X	(<i>D</i> 8) (<i>D</i> 17) (<i>D</i> 3) (<i>D</i> 1)

² The author has available for dissemination photostatic reproductions of the 32 stimuli.

Scoring

The standard and modified Rorschach responses were scored in the same manner for the following content categories.

Dynamic aggression. This category included responses which involved the ascription of aggressive-destructive activity or a state of death or injury to animals, objects, or people. The category consisted of three subcategories: (a) aggression—animals or people fighting, arguing, angry, or in a belligerent or aggressive posture, e.g., lions fighting, two ladies arguing, an octopus about to snatch a fish, a panther about to leap on its prey; (b) destruction—explosions or destructive action involving inanimate objects, e.g., bombs exploding, volcano erupting, A-bomb cloud, thrust of a rocket, landslide; the responses in this destruction category all involved destructive action (inanimate movement) and the response "rocket" or "bomb" alone was not included in this category; (c) injury—any animal, person, or object seen as dead, wounded, mutilated, or damaged, e.g., dead dog, squashed bug. This latter category did not occur too frequently and the first two categories (aggression and destruction) made up 75% of the dynamic aggression score.

Static aggression. This category included percepts of objects, animals, or animal parts which are normally regarded as aggressive but which did not involve the ascription of aggressive or destructive motivation or activity to the perceived object. This category included the following subcategories: (a) aggressive objects—objects or animal parts normally regarded as aggressive, e.g., bullet, bomb, volcano, rocket, knife, teeth, claws, horns; (b) aggressive animals—untamed animals normally regarded as aggressive, e.g., lions, tigers, panthers, bears, vampires; (c) monsters—creatures of myth or fantasy commonly regarded as aggressive, e.g., monster, giant, witch. Thus, the response "bomb" or "monster" was scored as static aggression, while the response "bomb exploding" or "monsters fighting" was scored dynamic aggression. Responses in which aggressive animals were engaged in nonaggressive activity were also scored static aggression, e.g., tigers walking, lions climbing, etc.

Dynamic and static aggression differ from each other in the degree to which the aggressive symbolism is disguised or attenuated. The dynamic aggressive response always involves undisguised aggressive action or the end result of aggressive-destructive activity. The aggressive implications of the static aggressive response are more disguised and symbolic. That is, the responses "knife" or "tiger" are less obviously aggressive in meaning than the responses "knife tearing flesh" or "tiger devouring a bird." This variable of "degree of disguise" was the primary basis upon which these two aggressive categories were separated.

Anatomy. Images of the internal body parts of animals or people or X rays of these internal body parts were scored anatomy, e.g., stomach, spinal column, X ray of the chest, lungs, kidney.

Dependence-orality. This category included imagery which many clinicians normally regard as indicative of dependency motivation. This variable included the following two subcategories: (a) passive-dependent activity—animals or people in a passive posture or in a dependent activity; e.g., people lying down, people resting or sleeping, pigs suckling, dogs nursing, people supporting each other; (b) oral content—responses in which people or animals were eating or drinking and responses describing food or objects directly associated with food or eating, e.g., people eating, people drinking, ice cream, fried chicken, people sitting in a restaurant, side of beef, pitcher of milk.

Human or human detail. This category included all images of humans or parts of humans, e.g., person, man sitting, face of someone, child, hands, etc.

Movement. All the Rorschach responses were coded separately for the standard categories of *M*, *FM*, and *m* as described by Klopfer et al. (1954). The present scoring of *m* includes Klopfer's *Fm*, *mF*, and *m*. A total score (*sum M*, *FM*, *m*) was obtained by summing the three movement scores.

Response length. The variable *R* referred to the number of spontaneous responses in each protocol.

The interrater reliability for the scoring of the first six categories was satisfactory for the over-all percentage of agreement between the author and an independent scorer was 92%.

RESULTS

Age and Sex Differences

Table 1 shows the frequency of occurrence of the variables for the three standard Rorschach protocols as well as the modified Rorschach for males and females separately.

There were no statistically significant age differences for the movement or content variables for the three standard Rorschach protocols. Occurrence of *sum M*, *FM*, *m* showed a moderate but nonsignificant increase with age for both sexes. The increase in response length from Protocol 1 to Protocol 3 was significant for males only ($p < .01$; two tails). Since the modified Rorschach elicited more responses than the standard plates and involved different stimuli, the differences in response frequencies between Protocol 3 and the modified Rorschach cannot be attributed solely to age changes.

None of the variables produced significant sex differences for all administrations. On Protocol 3, the males had a larger *sum M*, *FM*, *m* score ($p < .05$; two tails) but this was due, in part, to the fact that the males produced longer protocols than the females.

research had no a priori hypotheses about the outcome of the analysis, the data suggest several tentative conclusions. One of the purposes of the study was to investigate the long term stability of aggressive content. The differential stability and statistical independence of dynamic and static aggression indicate that these two variables are indices of different psychological processes. It is suggested, therefore, that future research with Rorschach aggressive content consider these two response classes separately. Research now in progress is directed at an explication of the psychological significance of the dynamic aggressive response.

As mentioned earlier, except for protocol length, movement responses showed the most consistent stability for the variables analyzed in this study. Seven of the 16 stability coefficients involving *M* and *FM* were statistically significant at the .05 level or better, and human movement had the highest average stability coefficient for the four interprotocol comparisons (average $\phi = .40$ for men and .29 for women). It is of interest to note that the investigations described earlier on the short term reliability of movement responses yielded coefficients that were not much higher than the ones reported here. Epstein et al. (1957) reported reliability coefficients for *M*, *FM*, and *sum M, FM, m* of .45, .49, and .61 respectively (all significant at $p < .01$) for protocols separated by three days. Kaplan and Berger (1956) reported correlations of .51 ($p < .01$) for *M* and .41 ($p < .05$) for *FM + m* with a four-day interval between tests. Thus the reliability coefficients for movement imagery with test retest intervals of several days did not differ markedly from the coefficients obtained in this study with an interval of several years. This finding indicates that the movement response, as now defined, is not reliable enough to be used for individual prediction. However, the present demonstration of stability over a three-year period for males and females and over a six-year period for males suggests that the human movement response is a crude index of a relatively stable personality construct.

Despite the fact that changes in values and motives occur during these years, it is generally assumed that some ego processes do re-

main stable over time. Research and theory suggest that the human movement response is associated with an introspective attitude and an awareness of motives, conflicts, and sources of anxiety. The research of Affleck and Mednick (1959), Gibby, Stotsky, Miller, and Hiler (1953), Singer, Wilensky, and McCraven (1956), and King (1958) indicate that movement responses are produced by individuals who tend to be aware of their motives, conflicts, and sources of anxiety and are willing to talk about these intrapsychic phenomena. These findings agree with the statements of Klopfer et al. (1954) that movement imagery reflects an availability to consciousness of thoughts that are directly related to basic motives and conflicts.

Current personality theory relies heavily on constructs which describe motives, defenses, or styles of behavior. The human movement response presumably taps a style of thinking or preferred manner of dealing with a motive rather than a motivational predisposition like dependency or aggression. Since human movement showed greater evidence of stability than most of the present content categories, it is possible that preferred modes of thinking (cognitive styles) may be more resistant to change than the strength of motives and values. Theoretical refinement of the movement variable together with the construction of superior instruments should produce more acceptable levels of reliability and a clearer understanding of the meaning of this behavior.

SUMMARY

This paper summarized data on the long term intra-individual stability of selected Rorschach categories for three standard Rorschach protocols and a modified set of Rorschach stimuli.

The Ss were 37 males and 38 females for whom standard Rorschach protocols were administered at median ages of $10\frac{1}{2}$, $13\frac{1}{2}$, and $16\frac{1}{2}$. In addition, 52 of these Ss were administered 32 modified Rorschach stimuli at a median age of 25 years. The modified stimuli were prepared from the standard plates and were congruent with common *D* and *d* areas. All protocols were scored for number of responses, human, animal, and inanimate move-

ment, sum of all movement responses, dynamic aggression, static aggression, anatomy, dependent-oral imagery, and human plus human detail responses.

The results revealed that:

(a) For the standard Rorschach protocols, number of responses, dynamic aggression, anatomy, dependence-orality, human movement, and sum movement showed low but statistically significant degrees of stability during the ages 13½ and 16½ (phi coefficients ranged from .39 to .59). Static aggression and human percepts showed no evidence of stability.

(b) In comparing the last standard Rorschach with the modified Rorschach stimuli, dynamic aggression and animal movement (for men) and human movement (for women) showed a significant degree of stability (phi coefficients of .60, .42, and .47 respectively).

(c) Dynamic aggression, which usually involved aggressive movement and contained undisguised aggressive content, was statistically independent of static aggression, which did not involve aggressive action and contained more disguised aggressive symbolism. In addition, dynamic aggression occurred more frequently to chromatic cards while static aggression was elicited more often by achromatic stimuli. It was suggested that future scoring codes for aggressive content differentiate between dynamic and static aggressive responses.

(d) Although the stability coefficients for human movement responses were low, this variable showed the most consistent stability during adolescence for the males. It was suggested that human movement responses reflected an "awareness of motives and conflicts" and that this construct appeared to be a relatively stable personality variable.

REFERENCES

- AFFLECK, D. C., & MEDNICK, S. A. The use of the Rorschach test in the prediction of the abrupt terminator in individual psychotherapy. *J. consult. Psychol.*, 1959, 23, 125-128.
- AMES, LOUISE B., LEARNED, JANET, METRAUX, RUTH W., & WALKER, R. N. *Child Rorschach responses*. New York: Hoeber, 1952.
- BAUGHMAN, E. E., & GUSKIN, S. Sex differences on the Rorschach. *J. consult. Psychol.*, 1958, 22, 400-401.
- DEVOS, G. A quantitative approach to affective symbolism in Rorschach responses. *J. proj. Tech.*, 1952, 16, 133-150.
- ELIZUR, A. Content analysis of the Rorschach with regard to anxiety and hostility. *Ror. Res. Exch.*, 1949, 13, 247-284.
- EPSTEIN, S., NELSON, JANE V., & TANOFKY, R. Responses to inkblots as measures of individual differences. *J. consult. Psychol.*, 1957, 21, 211-215.
- FINNEY, B. C. Rorschach test correlates of assaultive behavior. *J. proj. Tech.*, 1955, 19, 6-16.
- FISKE, D. W., & BAUGHMAN, E. E. Relationships between Rorschach scoring categories and the total number of responses. *J. abnorm. soc. Psychol.*, 1953, 48, 25-32.
- GIBBY, R. G., STOTSKY, B. A., MILLER, D. R., & HILER, E. W. Prediction of duration of therapy from the Rorschach test. *J. consult. Psychol.*, 1953, 17, 348-354.
- KAGAN, J., SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Personality and IQ change. *J. abnorm. soc. Psychol.*, 1958, 56, 261-266.
- KAPLAN, B., & BERGER, S. Increments and consistency of performance in four repeated Rorschach administrations. *J. proj. Tech.*, 1956, 20, 304-309.
- KING, G. F. A theoretical and experimental consideration of the Rorschach human movement response. *Psychol. Monogr.*, 1958, 72 (5, Whole No. 458).
- KLOPPER, B., AINSWORTH, MARY D., KLOPPER, W. G., & HOLT, R. R. *Developments in the Rorschach technique*. Vol. I. New York: World Book, 1954.
- PHILLIPS, L., & SMITH, J. G. *Rorschach interpretation: Advanced technique*. New York: Grune & Stratton, 1953.
- PIOTROWSKI, Z. *Perceptanalysis*. New York: Macmillan, 1958.
- RADER, G. E. The prediction of overt aggressive verbal behavior from Rorschach content. *J. proj. Tech.*, 1957, 21, 294-306.
- RANZONI, JANE H., GRANT, MARGUERITE Q., & IVES, VIRGINIA. Rorschach card pull in a normal adolescent population. *J. proj. Tech.*, 1950, 14, 107-133.
- SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing: Theory and application*. New York: Grune & Stratton, 1954.
- SINGER, J. C., WILENSKY, H., & MCCRAVEN, V. G. Delaying capacity, fantasy, and planning ability: A factorial study of some basic ego functions. *J. consult. Psychol.*, 1956, 20, 375-383.
- STAVRIANOS, BERTHA. An investigation of sex differences in children as revealed by the Rorschach method. *Ror. Res. Exch.*, 1942, 6, 168-175.
- TOWBIN, A. P. Hostility in Rorschach content and overt aggressive behavior. Unpublished doctoral dissertation, Yale Univer., 1955.
- WOLF, I. Hostile acting out and Rorschach test content. *J. proj. Tech.*, 1957, 21, 414-419.
- ZUBIN, J., ERON, L. D., & SULTAN, FLORENCE. A psychiatric evaluation of the Rorschach experiment. *Amer. J. Orthopsychiat.*, 1956, 26, 773-782.

(Received February 11, 1959)

MMPI FACTORS OF THE HARRIS AND THE WIENER SUBSCALES¹

JAMES C. LINGOES

Michigan State University

There are essentially two main points of view regarding the factor structure of the MMPI (Hathaway & McKinley, 1951). On the one hand, there are those like Wiener (1948) and Wheeler, Little, and Lehner (1951) who stress what might be called the two-factor position. The former has classified the items of the MMPI clinical scales into the two categories of obvious and subtle, while the latter have concluded on the basis of a factor analysis of the MMPI scales that the test, "... permits diagnosis mainly in terms of 'neurotic' or 'psychotic,' but not in terms of type of neurosis or psychosis or other more specific category" (Wheeler et al., 1951, p. 140). On the other hand, there are those like Harris and Lingo (1955) and Fisher (1957) who believe that there are reliably more than two dimensions to this test (the multifactor position).

It is the purpose of this study to test the relative merits of these two positions. Some considerations that might be advanced for the multifactor viewpoint are: (a) Since the original publication of the MMPI a very large number of special, cluster, factor, and rationally derived scales have been published.

¹ The major part of the statistical analysis was carried out while the investigator was on the staff of the Langley Porter Neuropsychiatric Institute, San Francisco. Appreciation is expressed to the University of California Computer Center at Berkeley for granting free use of the IBM 701 computer, and to the National Science Foundation for their support of basic research using this computer. Further appreciation is expressed to the Michigan State University Computer Center for free access to the MISTIC computer to complete the analysis. Grateful acknowledgment is expressed to C. F. Wrigley, R. E. Harris, H. H. Anderson, L. L. McQuitty, C. Hanley, and G. H. Crook for their helpful suggestions and criticisms in the preparation of this paper.

Some of these new scales have proved helpful in both clinical and nonclinical settings, e.g., Drake's Social I-E scale (1946), Feldman's Prognosis for Shock scale (1951), Barron's Ego Strength scale (1953), Gough's Dissimulation scale (1954), and Harris's Subscales (1955), among others. Yet, however useful some of these scales may be, it is improbable that there are as many dimensions of personality or functioning as there are existing scales. It seems equally unlikely, on the other hand, that as many scales as this could have been developed in such a variety of settings, on quite heterogeneous groups, for ostensibly different purposes, and using such seemingly different criteria, and yet hold the view that the MMPI is capable of measuring only two basic factors. (b) Clinicians in practical settings seem to operate under the assumption, in writing their MMPI interpretations, that the test is useful in tapping attitudes and traits as well as in separating patients from normals and they attempt to rely upon a configural approach in doing this. (c) Previous factor analytic results may have been due in part to the selection of factorially complex, composite scales which may have tended to obscure rather than clarify the factor structure of this test. A hypothetical example of how such a situation might arise is the following. Let us suppose that different factors were involved in the skills of multiplication, division, addition, and subtraction. Now it is conceivable, depending upon the tests included for analysis, that a different factor structure would result according to whether one used a total score based on these four tests or used subtest scores. One might in the former instance come out with a general quantitative factor and in the latter instance

with both a quantitative factor and common factors for each of the skills involved, since the variance of the separate tests would be allowed to vary independently from the total test score.

To test the multidimensional hypothesis, the investigator felt that the independently developed rational subscales of Harris (1955) and of Wiener (1948), to be described below, would provide an excellent opportunity, since these subscales represent two different hypotheses about the MMPI structure. Both investigators started with the standard MMPI clinical scales but classified the items in these scales differently.

DESCRIPTION OF MMPI SUBSCALES FOR ANALYSIS

The Wiener subscales were derived by sorting items from each clinical scale into two groups. Those items which were relatively easy to detect as indicating emotional disturbance were called the *Obvious* items, while the remaining items were called *Subtle*. There were two main criteria: (a) how frequently the item was scored in the given direction by the normative samples and (b) subjective estimation of the implied emotional disturbance of the item. Items which were infrequently answered by normals were automatically considered to be obvious, e.g., all items appearing on the F scale.

Since the Harris subscales have not yet been published, the rationale underlying their construction will be given in more detail. These subscales were derived in conjunction with a study of prediction of response to psychotherapy (Harris & Christiansen, 1946). Differences on the *Pd*, *Pa*, *Sc*, and *Ma* standard MMPI scales were obtained between groups responding well and poorly to therapy. Neither group, however, was very high on any of these scales. Although one might have inferred something like "subclinical psychotic trends" in those responding poorly, there remained the possibility of achieving a more psychological definition of the differences by an inspection of differences on the items within these scales. Subscales were formed by grouping together the items within each scale which seemed similar in content or which seemed to

reflect a single attitude or trait, i.e., item clusters were subjectively estimated. Upon scoring the two therapy groups on the various subscales, a number of differences emerged which led to an interpretation in terms of ego strength. Later research by Feldman (1951) and Barron (1953) on patients responding well and poorly to various kinds of somatic and psychotherapy suggested that this interpretation was a sensible one. Furthermore, the clusterings that Harris made reveal a coincidence of overlap with Comrey's analyses of the MMPI scales at the item level, notably on *Hy* (Comrey, 1957b) and on *Pa* (Comrey, 1958b), thus substantiating to a considerable extent the adequacy of Harris's estimates.

Eight of Wiener's scales (*HyO* and *HyS* being omitted because of their near perfect correlations with the Admittance and Denial cluster scales of Little and Fisher (1952), which are to be subsequently analyzed), and 28 of Harris's subscales (1955) constituted the 36 variables of this study. The capacity of the IBM 701 program would allow for no more than this number; hence, some selection was necessary. This selection, however, in no way restricts testing the multidimensional hypothesis. Table 1, containing the various subscale names and descriptions, gives some idea of the content of these scales.

HYPOTHESES

There would appear to be only four possible outcomes to a factor analysis of the Harris and the Wiener subscales (if one excludes the unifactorial position), i.e., (a) two major factors corresponding to the psychotic and neurotic factors of Wheeler et al. (1951); (b) one large factor with high loadings on the Obvious scales and another large factor with high loadings on the Subtle scales, which would support the two-factor view of Wiener (1948); (c) six factors, one for each of the six standard scales, with subscales derived from the same standard scale appearing on the same factor, e.g., Paranoia-Obvious, Paranoia-Subtle, and Harris's Paranoia subscales of *Pa*₁, *Pa*₂, and *Pa*₃, which would imply a scale homogeneity which the authors of the MMPI did not purposefully attempt to

achieve; and (d) a classification cutting across both the standard scales and the Wiener subscales, in conformity with the viewpoint of Harris and Lingoes (1955). Hypothesis 3 would not appear to be a serious contender, since various investigators have demonstrated the factorial complexity of the

standard scales, e.g., it is known that there are two quite different kinds of items in the Hysteria scale and that they are negatively correlated (Little & Fisher, 1952). The fourth hypothesis is favored by this investigator because of the considerations advanced at the beginning of this paper.

TABLE 1
SUBSCALE DESCRIPTIONS

Scale	Description	Scale	Description
<i>DO</i>	Depression-Obvious	<i>Pa₂</i>	Poignancy: thinking of oneself as something special and different from other people; high-strung; cherishing of sensitive feelings; overly subjective; thin-skinned
<i>DS</i>	Depression-Subtle	<i>Pa₃</i>	Affirmation of Moral Virtue: excessive generosity about the motives of others; righteousness about ethical matters; obtuse naïveté; denial of distrust and hostility
<i>D₁</i>	Subjective Depression: a negation of joy in doing things; pessimism, poor morale, and low self-esteem; complaints about psychological inertia and lack of energy for coping with problems	<i>Sc_{1A}</i>	Social Alienation: a feeling of lack of rapport with other people; withdrawal from meaningful relationships
<i>D₂</i>	Psychomotor Retardation: nonparticipation in social relations; immobilization	<i>Sc_{1B}</i>	Emotional Alienation: a feeling of lack of rapport with oneself; experiencing the self as strange; flattening or distortion of affect; apathy
<i>D₃</i>	Complaints about Physical Malfunctioning: preoccupation with oneself	<i>Sc_{2A}</i>	Lack of Ego Mastery, Cognitive: the admission of autonomous thought processes, strange and puzzling ideas
<i>D₄</i>	Mental Dullness: unresponsiveness; distrust of one's own psychological functioning	<i>Sc_{2B}</i>	Lack of Ego Mastery, Conative: feelings of "psychological weakness"; abulia, inertia, massive inhibition; regression
<i>D₅</i>	Brooding: ruminativeness; irritability	<i>Sc_{2C}</i>	Lack of Ego Mastery, Defect of Inhibition and Control: a feeling of not being in control of one's impulses, which may be experienced as strange and alien; at the mercy of impulse and feelings; dissociation of affect
<i>Hy₁</i>	Denial of Social Anxiety: social introversion	<i>Sc₃</i>	Sensorimotor Dissociation: a feeling of change in the perception of the self and the body image; feelings of depersonalization and estrangement
<i>Hy₂</i>	Need for Affection and Reinforcement from Others: implied in an (obtuse) denial of a critical or resentful attitude toward other people; impunitiveness; overly protested faith and optimism in other people	<i>MaO</i>	Hypomania-Obvious
<i>Hy₃</i>	Lassitude-Malaise: complaints about functioning below par physically and mentally; effortful keeping up of a good front; need for attention and reassurance	<i>MaS</i>	Hypomania-Subtle
<i>Hy₄</i>	Somatic Complaints: of a kind that suggest repression and conversion of affect	<i>Ma₁</i>	Amorality: a callousness about one's own motives and ends and those of other people; disarming frankness; denial of guilt
<i>Hy₅</i>	Inhibition of Aggression: expressed by concurrence with others; disavowal of violence	<i>Ma₂</i>	Psychomotor Acceleration: hyperactivity, lability; flight from "inner life" and anxiety; pressure for action
<i>PdO</i>	Psychopathic Deviate-Obvious	<i>Ma₃</i>	Imperturbability: affirmations of confidence in social situations; denial of sensitivity; proclamation of independence from the opinions of other people
<i>PdS</i>	Psychopathic Deviate-Subtle	<i>Ma₄</i>	Ego Inflation: feelings of self-importance to the point of unrealistic grandiosity
<i>Pd₁</i>	Familial Discord: struggle against familial control		
<i>Pd₂</i>	Authority Conflict: resentment of societal demands and conventions and parental standards		
<i>Pd₃</i>	Social Imperturbability: denial of social anxiety; blandness; denial of dependency needs		
<i>Pd_{4A}</i>	Social Alienation: feelings of isolation from other people; lack of belongingness; externalization of blame for difficulties; lack of gratification in social relations		
<i>Pd_{4B}</i>	Self-alienation: lack of self-integration; avowal of guilt, exhibitionistically stated; despondency		
<i>PaO</i>	Paranoia-Obvious		
<i>PaS</i>	Paranoia-Subtle		
<i>Pa₁</i>	Ideas of External Influence: externalization of blame for one's problems, frustrations, failures; in the extreme degree, persecutory ideas; projection of responsibility for negative feelings		

METHOD

Subjects

The four samples of Ss used in this study consisted of 150 male and 100 female neuropsychiatric patients (58% were outpatients and 42% inpatients), and 100 male and 100 female community Ss. These four groups will be hereinafter referred to by the abbreviations: Pm, Pf, Cm, and Cf, respectively. Median ages and age ranges of these four groups are: 32, 9-65; 33, 14-63; 25, 19-47; and 29, 19-62, years, respectively. Patient Ss were randomly chosen from the Langley Porter Neuropsychiatric Institute MMPI files, the only criteria for selection being that not more than 30 items were left unanswered on the test and that the patient had been tested on the group form. Community Ss were for the most part drawn from the staff at the same institution on a voluntary basis, while the remainder were drawn from cases used in another study (Lingoes, 1957) and from college students at the University of California at Berkeley,² using the same criteria for inclusion. No attempt was made to match these groups on various variables nor was there any attempt to make the groups homogeneous. The implications of this lack of matching and the heterogeneous nature of the samples will be discussed later. Suffice it to say here that the patient samples did not differ appreciably from each other in age or education. Both patient samples were in general older and of lower socioeconomic status than the community Ss.

Procedure

An independent factor analysis, employing Hotelling's (1933) principal components method, was applied to each set of correlations for the four groups of Ss. The squared multiple correlation (SMC) was used in the diagonal. This index is known to be the best universal lower bound for the communality (Guttman, 1940) and is calculated by taking each variable in turn as the dependent variable and computing the multiple correlation of that variable with the $p-1$ remaining variables and taking its square root. The SMC is thus seen to be a measure of predictable or observed common test variance, i.e., variance common to the test itself and the remaining $p-1$ tests. In an empirical investigation, using 15 communality estimates, Wrigley (in press) has found SMCs to be one of the best estimates of communalities in terms of stability, rate of convergence to, and discrepancy from, the true communalities. In discussing the logical and practical issues of the common-specific variance distinction in factor analysis and the problem of communalities, Wrigley (1957, p. 94) has offered a number of cogent arguments favoring the use of SMCs over that of communalities, e.g., "The problem of alternative sets of diagonal values satisfying the theory then disappears. SMC's are probably less influenced by increasing the sample of

persons, since the revised values, though more accurate, may be either higher or lower. Determining the diagonal entries becomes separate from determining the number of significant factors. One disadvantage is the lack of a significance test . . .," which has subsequently been provided by Tryon³ and used in this study to determine whether the factors rotated were significant.

A principal axes solution yields as many factors as there are tests. Using SMCs in the leading diagonal instead of unities, however, gives rise to factors having negative latent roots. Only those factors with positive latent roots and with at least one significant loading, i.e., P less than .05, as determined by Tryon's (1956) method were rotated. The method of rotation adopted was that of Kaiser (1958), called the Varimax method, an analytic solution approaching orthogonal simple structure; i.e., the number of zero loadings for any particular factor are maximized so that the factor variance is concentrated so far as possible in a few tests. The method operates on all possible pairs of factors and finds that angular transformation which maximizes the variance of the squared loadings, summed over the entire matrix, proceeding iteratively until convergence occurs. In order to preserve objectivity and make for comparability, no attempt was made to improve upon the rotations by hand methods, although it was possible that a better approximation to simple structure (using other criteria of Thurstone) might thereby have been achieved.

Reproducibility of the factors was investigated by using the Burt-Tucker⁴ (Burt, 1948, p. 185; Tucker, 1951) index of factor similarity. To provide some psychological content for the factors, items from the scales defining the factors were submitted to 10 psychologists with MMPI test experience for inspection and written comments as to what they thought the factors might be measuring.

³ One can determine the significance of any factor loading h_{ji} on any dimension or Factor C_i for the k dimensions retained, by computing the standard error of h_{ji} using the following formula:

$$\sigma h_{ji} = \sqrt{\frac{1 - h_i^2}{N - (k + 1)}};$$

where h_i^2 is the communality of the variable in question and N is the sample size. Thus if the factor loading lies outside the range $\pm \sigma 1.96 h_{ji}$, then the population value could not be zero at the 95% interval of confidence.

⁴ Burt and Tucker independently conceptualized one possible solution to the problem of factor similarity as involving an index of proportionality among the factor profiles. These indices range from +1.00 to -1.00 and can be interpreted much as r^2 in accounting for percentage of variance. Accordingly, one must obtain an index greater than .707 to insure that one factor is more like another than it is unlike that factor.

² Gratitude is expressed to Alex Sheriffs for the use of his MMPI files.

RESULTS

Number of Factors

Fifteen factors accounted for 98% of the common factor variance in the two patient groups and for 96% of the variance in the two community groups.⁵ The number of factors extracted and the manner in which they cut across the classifications of Wiener, Harris, and Hathaway and McKinley would strongly suggest that the MMPI is indeed factorially complex. It took five factors in each of the patient groups and seven in each of the community groups to account for only 80% of the variance (based on the unrotated factor matrices). The first factor in each of the analyses accounted for approximately 51, 46, 37, and 39% of the observed variance in the respective unrotated matrices. These results would seem to suggest the untenability of the two-factor position.

Factor Replication

Seven factors were reproduced, i.e., had indices of factor similarity of at least .70, in all four analyses, four factors were replicated in three, and three factors appeared in two. Patients were more like one another than they were like community Ss (12 factors in common), and likewise there was greater congruence among the community Ss (10 factors in common); the same thing might be said for the intrasex comparisons (there were 10 factors in common within each sex group). Agreement in factors is greater in this study than heretofore reported in the MMPI literature. This agreement was probably to a considerable extent a function of objectivity in computation, but to some degree it might also be partly spurious, since there exists the possibility that the amount of item overlap not only increased the size of the correlations, but may have also introduced correlation among factors across groups.

⁵ Tables 2-7, containing indices of factor similarity (Table 2), factor names and variances (Table 3), means and standard deviations (Table 4), scale intercorrelations (Table 5), unrotated factor loadings (Table 6), and Varimax rotated factor loadings (Table 7) for each of the four groups of Ss, have been deposited with the American Documentation Institute. Order Document No. 6119, remitting \$1.75 for microfilm or \$2.50 for photocopies.

Order of Factors

To facilitate presentation, the order of factors was determined first by consideration of factor stability and second by the relative sizes of the factors. In accord with this ordering, new factor numbers were assigned, e.g., Factor I (which differs from the first factor of any particular study by this notation). The size of each factor was determined as follows: based on the sum of the squared factor loadings summed across groups (when replicated), the percentage of the total observed common variance of the four groups was obtained, e.g., the variance of the four groups (based on the SMCs) was 120.26, while the total variance of the four groups on the first factor in each analysis (based on the sum of the squared factor loadings) was 30.51, and therefore Factor I accounted for approximately 25% of the variance.

Although there were both interdiagnostic and intersex differences, these were found to be for the most part a function of the size of the various factor loadings rather than major differences in the relative magnitudes or patterns of the loadings. Consistent differences across diagnostic groups and sexes will be mentioned when they occur, but only those factors will be presented which were replicated, based on the logic that greater confidence can be placed in them. Each of the factors will be related to the series of factor studies carried out by Comrey (1957a, 1957b, 1958a, 1958b, 1958c) and Comrey and Marggraff (1958) on the items of the MMPI clinical scales. Since there is an overlap between the items for Comrey's factors and the items constituting the Harris subscales, the opportunity for objectively but indirectly assessing similarity across factors and studies becomes possible.

Manner of Presentation of Results

Results are presented in the following order for factors obtained in all four analyses: (a) the factor's number and name; (b) the most highly loaded subscales on the factor with their loadings for the four groups, i.e., Pm, Pf, Cm, and Cf; (c) any additional subscales which have loadings greater than .39 for at least two analyses; (d) the indices of factor similarity; and (e) judges' descriptions

of the factor based on (b) above. For factors which were replicated in at least three analyses, Items (a)–(d) above are given with an indication of the groups included appearing alongside of the factor name.

Factors Found in All Four Analyses

Factor I: General Maladjustment. There were seven scales with high loadings in all four groups:

DO	Depression-Obvious (.88, .89, .83, .83)
D ₁	Subjective Depression (.92, .92, .84, .85)
D ₄	Mental Dullness (.90, .90, .84, .82)
D ₆	Brooding (.85, .83, .73, .74)
Hy ₃	Lassitude-Malaise (.85, .90, .61, .57)
Pd _{4B}	Self-Alienation (.83, .68, .69, .49)
Sc _{2B}	Lack of Ego Mastery, Conative (.89, .85, .71, .58)

Other subscales with loadings greater than .39 for at least two samples were:

D ₂	Psychomotor Retardation (.63, .62, .32, .31)
D ₃	Physical Malfunctioning (.57, .61, .24, .26)
Hy ₄	Somatic Complaints (.46, .43, .27, .26)
Pd _{1A}	Social Alienation (.57, .41, .45, .31)
Pd ₂	Poignancy (.48, .46, .22, .36)
Sc _{1B}	Emotional Alienation (.75, .72, .09, .23)
Sc _{2A}	Lack of Ego Mastery, Cognitive (.78, .71, .43, .37)

The Pm factor best exemplified Factor I. Indices of agreement with the remaining three groups were: .99, .95, .94.

The 10 judges reviewing the 69 items from the seven scales with the highest loadings (overlapping items were offered only once) agreed in describing this factor as one of generalized anxiety, depression, guilt, pessimism, and anxious worrying. It is much like the first factor found in a number of other investigations and may appropriately be called the General Maladjustment Factor or as Kassebaum, Couch, and Slater (1959) prefer to call it, a factor of Ego Weakness vs. Ego Strength. This factor includes the majority of the items from Comrey's (1957a, 1957b, 1958c) first factors for the *D*, *Hy*, and *Pd* scales and his second factor of the *Sc* scale (Comrey & Marggraff, 1958).

Factor II: Denial of Social Anxiety. Five scales had high loadings in all four groups:

Hy ₁	Denial of Social Anxiety (.82, .87, .81, .87)
Pd _S	Psychopathic Deviate-Subtle (.69, .71, .70, .69)
Pd ₃	Social Imperturbability (.73, .83, .83, .87)
Ma _S	Hypomania-Subtle (.39, .57, .55, .38)
Ma ₃	Imperturbability (.62, .80, .63, .77)

No other subscales had loadings greater than .39 in any sample. This factor was also best represented by the Pm group. Indices of agreement with the other three groups were: .94, .95, and .93.

The 47 items comprising the above five scales elicited such descriptions as: social ease, interpersonal dominance, extraversion, and social adaptability (middle scorers) and veiled aggression, social obtuseness, denial of social anxiety, and naïveté (high scorers). Both Factors I and II had items referring to anxiety, but in contrast to General Maladjustment, the items in the present factor bear upon inter- rather than intrapersonal variables. This factor includes the majority of items from Comrey's (1957b, 1958a, 1958c) second factor of the *Hy* scale, fourth factor of the *Pd* scale, and the first factor of the *Ma* scale. It is also of interest to note that this factor bears some resemblance to Kassebaum et al.'s (1959) second factor of Introversion-Extraversion.

Factor III: Loss of Control. Four scales had high loadings:

Sc _{2C}	Lack of Ego Mastery, Defect of Inhibition and Control (.55, .71, .52, .73)
Sc ₃	Sensorimotor Dissociation (.37, .50, .11, .75)
Ma _O	Hypomania-Obvious (.63, .73, .45, .65)
Ma ₂	Psychomotor Acceleration (.56, .85, .66, .55)

Only one other subscale had loadings greater than .39 in at least two samples, i.e., *DS* Depression-Subtle (−.20, −.57, −.18, −.40). The Pm factor showed the highest indices of agreement with the other groups, i.e., .92, .88, and .93.

The consensus classified the 45 items from the first four scales into the areas of: loss of emotional and impulse control, restlessness, incipient ego disintegration, and possible body image distortion. This factor would appear to be measuring at the high end, the failure of neurotic defenses, and in the middle ranges, inadequate control mechanisms. The highest factor loadings tended to appear in the two female groups, which might indicate a higher saturation of this factor among females. There was little congruence between these subscales and Comrey's (1958c) and Comrey and Marggraff's (1958) factors of the *Ma* and *Sc* scales.

Factor IV: Denial of Distrust and Hostility. High loadings (greater than .39) were obtained on three variables:

- Hy₂* Need for Affection and Reinforcement from Others (.63, .59, .78, .58)
PaS Paranoia-Subtle (.90, .87, .93, .91)
Pa₂ Moral Virtue (.84, .92, .93, .89)

Indices of .96, .95, and .95 were obtained between the Cf factor and factors in the remaining groups. Some indirect support for this factor can be adduced from Comrey's (1957b, 1958b) analyses of the *Hy* and *Pa* scales, inasmuch as his third factors from both of these scales overlap in items with the *Hy₂* and *Pa₂* subscales.

Descriptions such as: naïveté, denial of distrust and hostility, dependence upon the good will of others, impunitiveness, tolerance, repressive and submissive behavior, were obtained from the 25 items constituting the above three scales. Both the present factor and Factor II involve repression, denial, and naïveté, but whereas the latter deals with interaction ease, the present factor was more specifically concerned with defenses against both hostile and aggressive impulses.

Factor V: Familial Discord. The following two scales had the highest loadings:

- Pd₁* Familial Discord (.76, .74, .68, .72)
Sc_{1A} Social Alienation (.46, .41, .50, .37)

One other scale had loadings greater than .39, i.e., *PdS* Psychopathic Deviate-Subtle (.26, .40, .14, .43). The Pf group had indices of .94, .84, and .94 with the three remaining groups.

This factor was described as measuring essentially intrafamilial conflict, unresolved dependency strivings, feelings of being mistreated, misunderstood, deprived, and isolated. There were 31 items appearing in the first two scales. Some agreement between the items of *Pd₁* and Comrey's (1958c) 12th factor on the Pd scale was evident, although Comrey and Marggraff's (1958) analysis of the *Sc* scale split the *Sc_{1A}* items among two factors.

Factor VI: Inhibition and Apathy. Two variables had high loadings on this factor:

- DS* Depression-Subtle (.51, .43, .69, .50)
D₂ Psychomotor Retardation (.46, .49, .60, .71)

No other subscales had loadings greater than .39 for at least two samples. The community *S* groups had higher loadings on this factor than the patients. The Cf factor yielded indices of .86, .83, and .85 with the other three factors.

The 28 items from these two scales were described in the following terms: inhibition, overcontrol, rigidity, obstinacy, feelings of apathy and inertia, emotional constriction, etc. *D₂* items were distributed on two of Comrey's (1957a) *D* factors. Although only two variables had sufficiently high loadings to define this factor, it is nonetheless a common factor according to the factor model employed in this study, i.e., using SMCs. However, one might prefer to call such factors as these *near specifics* (Wrigley, 1957). The psychological or statistical significance of such factors is, however, an issue apart from the common-specific distinction and such factors will be included if replicated, pending their further investigation.

Factor VII: Social Nonconformity. Only one scale had consistently high loadings in all four groups:

- Pd₂* Authority Conflict (.69, .41, .71, .71)

The second highest loadings for the two patient groups appeared on *Ma₁* Amorality; with loadings of .56 and .77 for males and females, respectively. In the community groups, there was no consistency. Since the Pm factor yielded the highest agreement indices with the other factors, i.e., .81, .81, and .82, and since the patients had a second scale comparable in loadings to the first (*Pd₂*), it was decided to consider this a patient factor, defined by both *Pd₂* and *Ma₁*.

General agreement was evident in the comments of the judges, e.g., rebelliousness, argumentativeness, immaturity, hostility directed specifically against authority, potential for acting out, nonconformity, assertion of independence, and egocentricity. Only 16 items from two scales represented this factor. *Pd₂* items did not cluster together in Comrey's (1958c) analysis of the *Pd* items.

Factors Found in Three Analyses

Factor VIII: Hostility-Alienation-Projection (*Pm*, *Pf*, and *Cf*). Two scales had high load-

ing in all three samples:

- Pd_{4A}* Social Alienation (.47, .53, .66)
Pa₁ Ideas of External Influence (.75, .89, .86)

Additional subscales with loadings greater than .39 were:

- PdO* Psychopathic Deviate-Obvious (.33, .47, .44)
PaO Paranoia-Obvious (.72, .83, .19)
Pa₂ Poignancy (.41, .42, .14)
Sc_{1A} Social Alienation (.51, .50, .31)

The *Pf* factor had indices of similarity of .95 and .84 with the other two factors. There were 29 items in the first two scales. This would appear to be primarily a patient factor.

Factor IX: Somatic Complaints (Pm, Pf, and Cm). There were only two scales with loadings greater than .39 in at least two samples:

- Hy₄* Somatic Complaints (.66, .74, .65)
Sc₃ Sensorimotor Dissociation (.46, .50, .78)

The *Pm* factor yielded indices of .89 and .71 with the other two factors. This too is probably a patient factor. There were 33 items in these two scales.

Factor X: Poignancy (Pf, Cm, and Cf). Only one scale had consistently high loadings:

- Pa₂* Poignancy (.67, .81, .73)

The *Cm* factor had indices of .80 and .89 with the two female factors. In the community groups, *PaO* had the next highest loadings (.59 and .71, respectively), while in the patient groups, this scale was associated with *Pa₁*, which appeared on Factor VIII. There were only 9 items in the above scale and this factor was considered to be a near specific.

Factor XI: Poor Health (Pm, Cm, and Cf). The three groups had one scale in common with high loadings:

- D₃* Physical Malfunctioning (.63, .58, .48)

The *Cm* factor had an index of .83 with the *Cf* factor, but only a borderline index of .70 with the *Pm* factor. There were 11 items in this scale. In addition to *D₃*, the community samples had loadings of .61 and .46 on *Hy₃* Lassitude-Malaise, and consequently this factor was considered to be a nonpatient factor.

Factors Found in Two Analyses

Three factors which accounted for 6½% of the variance collectively appeared too small

and doubtful to deserve anything more than a mention. Factor XII had high loadings on *MaO* and *Ma₁* for the community groups; Factor XIII had high loadings on *MaS* for the *Pm* and *Cf* groups; and Factor XIV had high loadings on *Ma₄* for the two patient groups.

DISCUSSION

The present study of six standard MMPI scales has demonstrated a dimensionalization which is inconsistent with the two-factor positions of both Wiener (1948) and Wheeler et al. (1951). Instead, the MMPI appears to have: (a) a complexity not coextensive with the presently constituted six scales, and (b) a greater simplicity than might be inferred from the 28 titles given to the Harris subscales (1955). Implications of the present findings will be discussed under the following topic headings: (a) the multidimensional potentials of the MMPI, and (b) the "obvious-subtle" distinction.

The Multidimensional Potentials of the MMPI

The term *potentials* has been chosen advisedly, since a test does not per se possess dimensionality, but only in relation to a specified population or set of operations which produce the test responses. The results of this paper would lead one to conclude that the MMPI can reliably measure considerably more than two dimensions, e.g., neuroticism or psychoticism.

Of the 15 factors extracted there are seven which have emerged from four analyses (accounting for approximately 64% of the total observed variance). There are a further four derived from three analyses, accounting for an additional 19% of the variance. Admittedly, the number of factors obtained is partly a function of the number of variables included and may be to some extent an artifact resulting from item overlap. Nevertheless, to conclude that because the first two or three factors account for roughly half the variance that the test is just measuring this or that aspect of personality, e.g., sickness or general maladjustment, or is influenced primarily by some response set, is not justified on the present evidence.

Despite the increased confidence in the factors resulting from replication in a number of heterogeneous samples, the logical possibility is here raised of committing what might be called the *identity error*, i.e., inferring identical factors from identical covariance patterns. It may well turn out that different groups respond in like manners to the same items for entirely different reasons, either because the items have different meanings to different individuals or that the items themselves are complex factorially and have differing validities. Scales are being developed based on the replicated factors to test this hypothesis on various groups of Ss.

The "Obvious-Subtle" Distinction

An inspection of the rotated Varimax loadings on the eight Wiener scales revealed the *O* scales were scattered over at least three factors in all four analyses. In general, the same findings applied to the *S* scales. Furthermore, as in the case of the *PaO* items, this scale was associated with different factors, i.e., with *Pa*₁ in the patient samples and with *Pa*₂ in the community samples. If the Obvious and Subtle scales from Hysteria had been added, *HyO* would have probably appeared on Factors I, IX, and XI, while *HyS* would probably have come out on Factors II and IV. Whatever else the Obvious-Subtle items might be measuring, Wiener's conclusions, based on the intercorrelations among these scales and other variables, do not fully explain the above results, i.e., "They (the *Obvious* and *Subtle* scales) seem to have the *same meaning* and to perform the *same functions* for each scale" (Wiener, 1948, p. 169).⁶ Wiener's contention would seem to be on firmer ground in respect to the *O* scales, since they do yield similar sign patterns across a large number of factors (they are positively interrelated). It would appear, however, that an analysis of his scales in the context of the Harris subscales makes previously considered specific variance now common. This additional source of common variance is, moreover, greater than that attributable to obviousness per se, i.e., content does seem to make an important difference.

⁶ Italics and parentheses present author's.

It would appear from the present evidence that obviousness and subtleness may have different as well as common meanings and functions within each scale and that the distinction Wiener has made, while useful in investigating response sets, must be accordingly modified to admit of many kinds of obvious and subtle items. This conclusion would appear to be supported by that of Hanley (1959) in respect to his investigation of acquiescence.

Conclusions

It can be said that this study has demonstrated the multidimensional potential of the MMPI and has lent indirect support to the configural approach to interpretation advanced by the authors of this test. Although item overlap may have introduced extra sources of variance, thus inflating the SMCs, it is felt that the subscale approach used here is a valuable one for clarifying factor structures. Because only six of the clinical scales were analyzed, the inclusion of other MMPI scales might have modified, clarified, or introduced factors to some undetermined degree. One might expect, however, that *Pt* would have appeared on Factor I and that *Hs* would probably have split on Factors III, IX, and XI had they been included.

It is believed that this study complements that of Comrey to the extent that some objective evidence can be adduced that the factors he obtained within scales are similar from scale to scale. This evidence is, of course, indirect, since while there is congruence between some of his factors and the Harris subscales, the correspondence is far from perfect. Factor scales to be developed based on this analysis are intended to supplement and not to replace the standard MMPI scales, although they might well replace the Harris subscales as an interpretative aid.

SUMMARY

This investigation was concerned with testing the merits of the two-factor versus the multifactor hypothesis of the MMPI as represented by Wiener and by Harris, respectively. Harris classified MMPI items under 28 categories, while Wiener chose only 2, i.e.,

Obvious and Subtle, as an adequate description of the test's complexity.

From the 36 subscales analyzed (representing six clinical scales) in four samples of subjects (patients and nonpatients, males and females), seven factors were replicated in four analyses and four in three. The 9 of these 11 factors which could be readily identified were: General Maladjustment (I), Denial of Social Anxiety (II), Loss of Control (III), Denial of Distrust and Hostility (IV), Familial Discord (V), Inhibition and Apathy (VI), Social Nonconformity (VII), Hostility-Alienation-Projection (VIII), and Somatic Complaints (IX).

It was concluded on the basis of the analyses that: (a) reliably more than two dimensions can be demonstrated; (b) the "obvious-subtle" distinction, while having some pertinence for the study of response sets, cannot fully explain the present factor results and that one must admit of many kinds of obviousness and subtleness; (c) while the present study is far from definitive, the fruitfulness of performing analyses on the basis of subscales rather than composite scales is suggested; and (d) evidence was offered relating the series of factor studies carried out by Comrey at the item level to the scale analysis of this study.

REFERENCES

- BARRON, F. X. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 27, 327-333.
- BURT, C. The factorial study of temperamental traits. *Brit. J. Psychol., Statist. Sect.*, 1948, 1, 178-203.
- COMREY, A. L. A factor analysis of items on the MMPI Depression scale. *Educ. psychol. Measmt.*, 1957, 17, 578-585. (a)
- COMREY, A. L. A factor analysis of items on the MMPI Hysteria scale. *Educ. psychol. Measmt.*, 1957, 17, 586-592. (b)
- COMREY, A. L. A factor analysis of items on the MMPI Hypomania scale. *Educ. psychol. Measmt.*, 1958, 18, 313-323. (a)
- COMREY, A. L. A factor analysis of items on the MMPI Paranoia scale. *Educ. psychol. Measmt.*, 1958, 18, 99-107. (b)
- COMREY, A. L. A factor analysis of items on the MMPI Psychopathic Deviate scale. *Educ. psychol. Measmt.*, 1958, 18, 91-98. (c)
- COMREY, A. L., & MARGGRAFF, W. M. A factor analysis of items on the MMPI Schizophrenia scale. *Educ. psychol. Measmt.*, 1958, 18, 301-311.
- DRAKE, L. E. A social IE scale for the MMPI. *J. appl. Psychol.*, 1946, 30, 51-54.
- FELDMAN, M. J. A prognostic scale for shock therapy. *Psychol. Monogr.*, 1951, 65 (10, Whole No. 327).
- FISHER, J. An empirical study of the relation of physical disease to body-object cathexis. Unpublished manuscript, VA Hospital, 42nd & Clement, San Francisco, Calif., 1957.
- GOUGH, H. G. Some common misconceptions about neuroticism. *J. consult. Psychol.*, 1954, 18, 287-292.
- GUTTMAN, L. Multiple rectilinear prediction and the resolution into components. *Psychometrika*, 1940, 5, 75-90.
- HANLEY, C. Responses to the wording of personality test items. *J. consult. Psychol.*, 1959, 23, 261-265.
- HARRIS, R. E., & CHRISTIANSEN, C. Prediction of response to brief psychotherapy. *J. Psychol.*, 1946, 21, 269-284.
- HARRIS, R. E., & LINGOES, J. C. Subscales for the MMPI: An aid to profile interpretation. Unpublished manuscript, The Langley Porter Neuropsychiatric Inst., 1st & Parnassus Aves., San Francisco 22, Calif., 1955.
- HATHAWAY, S. R., & MCKINLEY, J. C. *Minnesota Multiphasic Personality Inventory: Manual*. (rev.) New York: Psychological Corp., 1951.
- HOTELLING, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, 1933, 24, 417-441, 498-520.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.
- KASSEBAUM, G. G., COUCH, A. S., & SLATER, P. E. The factorial dimensions of the MMPI. *J. consult. Psychol.*, 1959, 23, 226-236.
- LINGOES, J. C. MMPI test correlates of Szondi picture preferences. *Szondi Newsltr.*, 1957, 6, No. 1.
- LITTLE, K., & FISHER, J. The internal structure of the MMPI Hysteria scale. Paper read at Bay Area Veterans Administration Clinical Psychology Research Meeting, San Francisco, Oct., 1952.
- TUCKER, L. R. A method for synthesis of factor analysis studies. *U. S. Depart. Army, Personnel Sec. Rep.*, 1951, No. 984.
- TRYON, R. C. Tests of significance and salience in cluster and factor analysis on principles of domain sampling. Unpublished manuscript, Univer. of California, Contract Report No. 15, 1956.
- WHEELER, W. M., LITTLE, K. B., & LEHNER, G. F. J. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, 15, 134-141.
- WIENER, D. N. Subtle and Obvious keys for the MMPI. *J. consult. Psychol.*, 1948, 12, 164-170.
- WRIGLEY, C. F. The distinction between common and specific variance in factor theory. *Brit. J. Statist. Psychol.*, 1957, 10, 81-98.
- WRIGLEY, C. F. An empirical comparison of various methods for the estimation of communalities. *Educ. psychol. Measmt.*, in press.

(Received February 16, 1959)

THE RESULTS OF PSYCHOTHEAPY WITH CHILDREN: A REVALUATION

J. HOOD-WILLIAMS

University of Natal, South Africa

Recently Levitt (1957) published a survey of the studies concerning the effectiveness of psychotherapy with children on the model of Eysenck's (1952) now classic enquiry with adults. Levitt concludes that the published figures do not support the contention that psychotherapy with children is more effective than nonintervention. His baseline data are derived from two studies: that by Witman and Keller (1942) and that by Lehrman et al. (1949). Averaging these studies shows that untreated controls have an improvement rate of 72.5%, whereas a survey of 18 studies of the outcome of psychotherapy with children shows an average improvement rate of 67.05%. More strictly comparable with the baseline data studies is the averaging by Levitt of 17 follow-up studies of the effects of psychotherapy on children. These yield a figure which shows that 78.22% of the cases were improved.

Levitt's article is a challenge to all workers in the field of child psychotherapy, indicating as it does the relative ineffectiveness of this field of endeavor. Levitt is careful to point out that it is *not proven* that psychotherapy is ineffective; but the converse, that psychotherapy *is* effective, is also not proven.

The present study was undertaken as an attempt to isolate from the published literature some of the factors contributing to success or failure in psychotherapy as rated at the close of treatment. Follow-up studies were excluded since criteria for follow-up success were not detailed in any definable way.

THE SOURCES OF UNCERTAINTY

The problem of establishing a baseline. At this juncture the writer wishes to note the difficulty of arriving at baseline data for the

evaluation of the effectiveness of psychotherapy. The method used in the two studies (Lehrman et al., 1949; Witman & Keller, 1942) accepted by Levitt has serious drawbacks. These studies used as controls cases who had made an approach to the clinic and had been put on the waiting list but who, for one reason or another, did not accept the offer of treatment when it was later made available. Levitt says that such cases were in all respects similar to treated cases, except for the fact that they received no treatment. While the published literature contains no studies where the controls have been obtained by any other method (so that no better statistics are available as a basis for comparison), it is, however, not entirely a valid method of obtaining controls.

For one thing, too little is known about the factors intervening between first clinical contact and later acceptance for treatment, and, for another, it is common experience at all clinics that many cases on the waiting list clear up before treatment can be given. Too little is known about the dynamics of this, but at least two factors are known to account for it: (a) cases where the approach to the clinic is made in a temporary crisis, which subsides while the case is pending, and (b) cases where a healing dynamic process is initiated by the parents' recognition of their need for help, so that by the time help is available it is no longer needed.

As all these factors are operative for both the treated and untreated groups, is it not possible that the treated groups represent the more serious, intractable problems, while at least a proportion of the untreated groups are minor or transient problems?

However, in the absence of any better sta-

tistics, these must be accepted, albeit with reservations, whose very nature demands that conclusions drawn from them should be highly tentative.

Comparison of the studies showing high success with those reporting low success. The four studies reporting the highest percentage of success (Cohen & Davis, 1934; Irgens, 1936; Newell, 1943; Witmer & others, 1933) were compared with the four studies reporting the lowest (Brown, 1947; Christianson, Gales, & Coleman, 1934; Cunningham, Westerman, & Fischhoff, 1956; Maas, Kahn, Stein, & Sumner, 1955).¹ This brought to light several features that were blurred in Levitt's (1957) treatment of the data.

The influence of the date of the study. Most unexpectedly, it was found that the cases treated in clinics in the high success studies all date from 1931–1932 despite the divergence of dates of publication; whereas three of the four low success studies date from after 1945. The one exception among the low success studies (Christianson et al., 1934), which also dates from 1931–1932, is differentiated from the other studies, which all deal with major clinics in major cities, by being concerned with the northern New Jersey clinics—i.e., the possibility must not be ignored that this study may reflect a lower level of service than the others.

The entire 18 studies were then arranged into two categories, dichotomized at the average level of success (67%), with the following findings as regards dates:

Highs (over 67%)

Author	Percentage	Date
Irgens	85.7	1936
Cohen and Davis	80.7	1934
Newell	79.2	1943
Witmer & others	79.0	1933
Lee and Kenworthy	78.6	1929
Burlingham	76.0	1931
Barbour	74.3	1951
Hubbard and Adams	73.0	1936
Reid and Hagan	72.8	1952
Carpenter	67.7	1939

Lows (under 67%)

Author	Percentage	Date
La More	66.0	1941
Lehrman et al.	65.3	1949
Albright and Gambrell	65.0	1938
Christianson, Gales, & Coleman	62.7	1934
Cunningham, Westerman, & Fischhoff	59.8	1956
Brown	59.3	1947
Canaday	47.8	1940
Maas, Kahn, Stein, & Sumner	43.1	1955

From immediate inspection it can be seen that, of the Highs, 7 of the 10 are studies prior to 1940. Of the studies dated later than 1940, one (Barbour, 1951) is a British study and not strictly comparable with the other United States studies; another (Newell, 1943), though published in 1943, is concerned with data gathered 10 years earlier, in 1933; and a third (Reid & Hagan, 1952) is related to residential treatment while all the others are concerned with nonresidential treatment.

Among the Lows, on the other hand, only two of the eight studies are dated 1940 or earlier. One of these (Christianson et al., 1934) has been discussed above and is not regarded as strictly comparable with the others in this group. The other (Canaday, 1940) has not been available for detailed study, for reasons given above.

Consequently, if the Barbour (1951) and the Reid and Hagan (1952) studies are omitted from the Highs for the reasons given, and the Christianson et al. (1934) study from the Lows, one finds that all the Highs date from before 1939, and six out of seven of the Lows date from after 1940.

The problem of length and intensity of treatment. It was remarkable how little information could be obtained about the important questions: (a) How much treatment did each case receive, and (b) Was the person concerned with the treatment of the child a qualified psychotherapist? ² This is

¹ One of the studies included by Levitt that reported the second lowest percentage improvement (Canaday, 1940) was excluded from this study as it is an unpublished thesis and therefore inaccessible to study at this distance. The four "lows" in this study are, accordingly, four of the five lowest included by Levitt.

² The writer is aware that the term "psychotherapist" is open to various divergent interpretations. It is here being used to refer to anyone having spe-

particularly curious in view of the vast body of information that was reported anent presenting symptoms, age, sex, socioeconomic status, parental attitude, and other background variables. Only one study that was reviewed (Maas et al., 1955) included a breakdown of the success-failure percentages in terms of length of treatment. This study, which reports the lowest over-all improvement rate (only 43.1%), shows that successful outcome is directly related to length of treatment.

No. of Interviews	N	Percentage
		Improved
None or single family contact	137	7.2
Brief service (2-9 interviews)	77	22.1
Short-term treatment (10-39 interviews)	144	53.8
Long-term treatment (40 or more interviews)	144	80.2

By contrast one finds that in the four studies reporting the highest level of success, the actual amount of treatment each child received was mostly extremely slight. Thus Irgens (1936) gives details of the treatment for 48 out of the 70 cases comprising the study. Of these, only four had more than 10 interviews, and these were with an unspecified worker. Her reports on treatment are characteristically along these lines: "Two interviews, summer camp, school adjustment." Yet this study has the highest success level (85.7%). Compare this with Newell (1943) who reports 79.2% success for a procedure designed specifically for use in a school system and dealing exclusively with difficulties arising in school, where the great preponderance of the cases (percentage unspecified) are seen once only, treatment being concentrated on advice to teachers about handling, and secondarily to parents. This is in sharp contrast with the later studies, notably the previously cited one by Maas et al. (1955), where the majority of cases (57.4%) had over 10 interviews each.

Related to the problem of length of therapy is, of course, that of the qualifications of the specialized training in child psychotherapy, irrespective of theoretical approach, and including psychiatrists, clinical psychologists, and child analysts, but excluding social workers, teachers, etc.

therapist. Once again there is a dearth of information, but in the only study which does specify the type of skills of the therapist, the resulting information is interesting. Witmer and students (1933), who report an overall improvement rate of 79.0%, also broke down the figures to indicate whether the therapist was a psychiatrist or a social worker. Her findings are not as dramatic as Maas', showing percentages varying from 73 to 82 for both psychiatrist and social worker for *less than 15* interviews; but at *over 15* interviews the psychiatrist achieves 97% success ($N = 30$), and the social worker 84% ($N = 208$).

The problems of the inclusion of unsuitable cases. It was noted that, where information was included about presenting symptoms, in every instance there were included small numbers of cases that are presumably either incurable (e.g., mongoloidism) or curable by somatic medicine (e.g., endocrine dysfunctioning). Though the numbers are never large (the percentages are never in excess of 10% of all cases), their exclusion slightly improves the percentages for successful treatment.

DISCUSSION

In a sense, Levitt (1957) has put the contemporary child guidance clinic to the test and found it wanting. But is this in fact so? Two main arguments suggest it is not. First, there are serious inadequacies (detailed above) in the data relating to the recovery rate of untreated children which render them suspect as a basis for comparison with treated cases. Secondly, the studies straddle a period in which child guidance clinics have not only been growing and modifying their practices but also adjusting to rapid social changes in the community as a whole, so that all other considerations apart (such as lack of uniformity of procedures in individual child guidance clinics), it is unsatisfactory to arrive at conclusions based on the amalgamation of these studies, for amalgamation treats them all as equal; whereas, as has just been suggested, it is more probable that a close examination will reveal that child guidance practice in, say, 1933, had little in common with that obtaining in 1950 or 1955. However, his study does show that *present-day child guidance clinic practices* are not highly

effective in dealing with children's psychological malfunctionings, though it does appear that they were considerably more effective in the early 1930's than they are now. Why should this be so?

No definite and final answer can be given here, but is it not possible (from the internal evidence of the various studies reviewed) that during the past three decades child guidance clinics have undergone a variety of changes both in the nature of the service rendered to the community and, even more importantly, in the severity of the disturbances they attempt to deal with?

It is suggested that during the past two decades, child guidance clinics have increasingly been dealing with deep-seated cases of disturbances requiring extensive psychotherapy and less with relatively minor cases that can be treated by manipulation of the environment.³ Probably there is a multiplicity of reasons for this change, including increasing public awareness of severe childhood disturbance, increasing availability of trained therapists, possibly also an increased actual incidence of more severe disturbance, reflecting the social dislocation in the past two decades, and possibly many other factors as well.

It is the writer's contention that the question "How effective is psychotherapy with children" is still unanswered.

SUMMARY

In an attempt to throw further light on Levitt's (1957) findings on the apparent ineffectiveness of psychotherapy with children, the unsatisfactoriness of the baseline data was discussed. The literature was re-reviewed, and the method of extreme groups was applied. The four studies reporting the highest percentages of success were contrasted with the four reporting the lowest. It was found that:

1. The studies reporting high percentages of success relate to cases treated during the years 1931-1932, whereas three of four stud-

³ This is borne out to some extent by the fact that early studies make no mention of psychotic children, while later ones do. Also, where treatment is specified, earlier studies rely heavily on environmental manipulation, while later studies report increasing numbers of interviews with the patient himself.

ies reporting low percentages are postwar studies. This relationship was found to hold good when a total of 18 studies were dichotomized at the average percentage.

2. There was little agreement among the studies regarding factors responsible for higher successes, except that length of treatment was related positively to successful outcome, as was also the nature of the skills of the therapist.

3. The wide divergences in criteria for success, and methods of treatment, render the averaging of a large number of studies an unsatisfactory method of evaluating the effectiveness of psychotherapy with children.

It is contended that, as yet, there is no satisfactory answer to the question of how effective psychotherapy is with children.

REFERENCES

- ALBRIGHT, S., & GAMBRELL, H. Personality traits as criteria for psychiatric treatment of adolescents. *Smith Coll. Stud. soc. Wk.*, 1938, 9, 1-26.
- BARBOUR, R. F. Selected surveys prepared for the interclinic conference. In J. F. Davidson (Chairman), *Follow-up on child guidance cases*. London: 1951 Ninth Child Guid. Inter-Clinic Conf., 1951. Pp. 49-59.
- BROWN, MARJORIE. Adolescents treatable by a family agency. *Smith Coll. Stud. soc. Wk.*, 1947, 18, 37-67.
- BURLINGHAM, S. A quantitative analysis of psychiatric social treatment carried out in seventy-five cases at the Institute for Juvenile Research. Unpublished master's thesis, Smith Coll., 1931.
- CANADAY, L. J. A way of predicting the probable outcome of treatment of young children who run away. Unpublished master's thesis, Smith Coll., 1940.
- CARPENTER, J. A. Some factors relating to the method and outcome of case-work treatment with the adolescent girl when the girl herself is the focus of treatment. Unpublished master's thesis, Smith Coll., 1939.
- CHRISTIANSON, E., GALES, M. E., & COLEMAN, F. A survey of the intake of a mental hygiene clinic, with special reference to the outcome of treatment. *Smith Coll. Stud. soc. Wk.*, 1934, 5, 211-212.
- COHEN, MARION, & DAVIS, ELLEN. Factors related to the outcome of treatment in a child guidance clinic. *Smith Coll. Stud. soc. Wk.*, 1934, 5, 212-214.
- CUNNINGHAM, J. M., WESTERMAN, H. H., & FISCHHOFF, J. A. A follow-up study of children seen in a psychiatric clinic for children. *Amer. J. Orthopsychiat.*, 1956, 26, 602-611.
- EYSENCK, H. J. The effects of psychotherapy: An evaluation. *J. consult. Psychol.*, 1952, 16, 319-324.
- HUBBARD, R. M., & ADAMS, C. F. Factors affecting the success of child guidance treatment. *Amer. J. Orthopsychiat.*, 1936, 6, 81-102.

- IRGENS, EFFIE M. Must parents' attitudes become modified in order to bring about adjustment in problem children? *Smith Coll. Stud. soc. Wk.*, 1936, 7, 17-45.
- LAMORE, MARY T. An evaluation of a state hospital child guidance clinic. *Smith Coll. Stud. soc. Wk.*, 1941, 12, 137-164.
- LEE, P. R., & KENWORTHY, M. E. *Mental hygiene and social work*. New York: Commonwealth Fund, 1929.
- LEHRMAN, L. J., et al. Success and failure of treatment of children in the child guidance clinics of the Jewish Board of Guardians. New York City Jewish Bd. Guard. Res. Monogr., 1949, No. 1.
- LEVITT, E. E. The results of psychotherapy with children: An evaluation. *J. consult. Psychol.*, 1957, 21, 189-196.
- MAAS, H. S., KAHN, A. T., STEIN, H. D., & SUMNER, DOROTHY. Socio-cultural factors in psychiatric clinic services for children. *Smith Coll. Stud. soc. Wk.*, 1955, 26, 1-90.
- NEWELL, H. W. The methods of child guidance adapted to a public-school program. *Ment. Hyg., N. Y.*, 1943, 18, 362-373.
- REID, J. H., & HAGAN, HELEN R. *Residential treatment of emotionally disturbed children*. New York: Child Welf. League of Amer., 1952.
- WITMAN, H. L., & KELLER, J. Outgrowing childhood problems: A study in the value of child guidance treatment. *Smith Coll. Stud. soc. Wk.*, 1942, 13, 74-90.
- WITMER, HELEN L., & others. The outcome of treatment in a child guidance clinic: A comparison and evaluation. *Smith Coll. Stud. soc. Wk.*, 1933, 3, 339-399.

(Received March 25, 1959)

REPLY TO HOOD-WILLIAMS

EUGENE E. LEVITT

Indiana University Medical Center

The commentary by Hood-Williams (1960) casts some interesting and thought provoking light on my review of the results of psychotherapy with children (Levitt, 1957b). Hood-Williams' contentions are certainly worthy of discussion, but they hardly constitute a heading so encompassing as "revaluation," as I shall attempt to demonstrate.

The influence of the date of the study. If one accepts Hood-Williams' analysis, its effect is to emphasize the inference that treatment has been ineffective. We must allow that the work since 1945 is more representative of what we *now* call psychotherapy than work done previous to that time. Hood-Williams suggests, as an explanation for the date factor, that clinics are now dealing with more serious problems than they used to. Possibly. It appears more likely, though, that the lower recent improvement rate is a function of advances in diagnostic methods, greater knowledge of personality dynamics, and increased research sophistication.

On the other hand, one should not hasten to accept Hood-Williams' analysis. In my Table 5 (Levitt, 1957b, p. 194), I pointed to a potential positive relationship between outcome and number of points used in the outcome rating scale. All of the reports using 4- and 5-point scales show above average improvement rates, and all but one were carried out prior to 1936. Both of the reports using dichotomous scales have below average improvement rates; they were reported in 1955. Thus there are two possible causal factors here.

The problem of establishing a baseline. I qualified the comparability of treated cases and so-called "defector" controls with the phrase that they were similar "so far as is known" (Levitt, 1957b, p. 190). I chose what appeared to me to be the best kind of con-

trol, and Hood-Williams is in error when he says that there have been no other methods published. At least five control techniques have been employed in the past (cf. Levitt, Beiser, & Robertson, 1959).

The evidence accumulated since the publication of my 1957 paper suggests that the idea that the treated cases are serious disturbances, while the defectors are transient disorders, is also erroneous. One study (Levitt, 1957a) shows that treated and defector child cases do not differ on some 61 factors, including two clinical estimates of severity of disturbance, and eight other factors relating to symptoms. A second study (Levitt, 1958a) reported that expert judges found no difference in severity of symptoms between treated and defector child cases, judging from case records. The mean severity ratings, based on a five-point scale, were 2.98 for treated cases and 3.02 for defectors.

Follow-up interviews with parents of 142 defector cases (Levitt, 1958b) disclosed that only 13% attributed defection to improvement in the child's symptoms. Over half blamed either the clinic or environmental circumstances. In an earlier telephone follow-up of 40 cases (Inman, 1956) 14, or 35%, gave improvement as a reason for defection. However, all but three of these parents saw fit to volunteer at least one other reason as well, which hints at some defensiveness about the improvement explanation.

Putting the two studies together, 18% of the parents gave improvement in the child as a reason for defection. This percentage does not seem large enough to warrant the statement that pretherapy remission distinguishes treated cases and defectors, though it may have affected the intergroup comparison to some extent. However, we have no independent evidence to suggest that the same propor-

tion would not hold for the *treated* cases as well. A parent might well feel that the child had improved and yet not see fit to break contact with the clinic.

I do not claim that any of these studies or lines of reasoning is conclusive, but taken together, their weight seems opposed to the hypothesis that the defector cases are less disturbed as a group than cases who go on to receive treatment.

The problem of length and intensity of treatment. It was not my intention to deal with these factors in my review, primarily because the needed information was lacking in most reports, as Hood-Williams notes. The problem is certainly noteworthy, especially since it is by no means as simple to grapple with as it might appear at first glance. Consider Maas, Kahn, Stein, & Sumner's (1955) table presented by Hood-Williams. If we accept these data as representative, and compare them with the defector results, the inference is not simply that "successful outcome is directly related to length of treatment." The data show that improvement is a function of at least 40 treatment interviews; a lesser number actually makes the patient *worse*. Would the practicing psychotherapist care to support the contention that 35 treatment interviews is worse than no treatment at all? Is there an hypothesis that could explain this peculiar phenomenon? And why is it that only 7% of the Maas' sample having no contact, or only one family contact, with the clinic showed improvement, while the figure is 10 times as great for approximately the same degree of contact among defectors?

Attempts to determine the relationship between outcome at close of treatment and number of interviews are confounded by a time factor. Twenty interviews usually span a longer time period than five interviews. Should a positive relationship appear, we cannot be certain that it is not due simply to the passage of time alone, quite apart from number of interviews. This point is exemplified by Table 4 in my review (Levitt, 1957b, p. 193) and by the Denker (1946) study of untreated adults. Both show increases in improvement rate with time, without reference to intensity of treatment.

I agree entirely with Hood-Williams that the matters of intensity and type of treatment, and the qualifications of therapists are potentially significant, and should be thoroughly discussed in any evaluation study. (In the review, I referred to the "generally poor caliber of methodology and analysis" in the studies covered therein.) The fact that this kind of information is usually lacking suggests that one should be circumspect in drawing inferences from the review. It does not impeach the findings.

The character of the evaluated treatment is taken up in some detail in the evaluation follow-up study at the Institute for Juvenile Research in Chicago (Levitt et al., 1959). We report that almost half of the cases were treated by students or staff members with one year of experience or less. Only one-third were seen by therapists with more than three years of experience. A third had to switch therapists at least once during the course of therapy, and some had been handled by three or even four therapists. Ninety per cent of the individuals were treated on a once-a-week basis, and only 15% by psychiatrists.

If similar conditions of therapy have been prevalent among child guidance clinics, then Hood-Williams' point might have greater import. But we do not know, and probably will never know. In fact, we do not even know for certain that experienced therapists will be more effective, though it seems to be a plausible hypothesis.

Hood-Williams has raised some reasonable points. If his paper serves to focus additional attention on the need for continuing evaluation of our therapeutic efforts, and on the necessity for tightening up methodology and technique, then its publication is amply justified. But his conclusions concerning the review of results are too sweeping and incautious for the limited and equivocal analyses which he carried out.

REFERENCES

- DENKER, P. G. Results of treatment of psychoneuroses by the general practitioner. *N. Y. State med. J.*, 1946, 46, 2164-2166.
- HOOD-WILLIAMS, J. The results of psychotherapy with children: A reevaluation. *J. consult. Psychol.*, 1960, 24, 83-87.

- INMAN, ANN C. Attrition in child guidance: A telephone follow-up study. *Smith Coll. Stud. soc. Wk.*, 1956, 27, 34-73.
- LEVITT, E. E. A comparison of "remainers" and "defectors" among child clinic patients. *J. consult. Psychol.*, 1957, 21, 316. (a)
- LEVITT, E. E. The results of psychotherapy with children: An evaluation. *J. consult. Psychol.*, 1957, 21, 189-196. (b)
- LEVITT, E. E. A comparative judgmental study of "defection" from treatment at a child guidance clinic. *J. clin. Psychol.*, 1958, 14, 429-432. (a)
- LEVITT, E. E. Parents' reasons for defection from treatment at a child guidance clinic. *Ment. Hyg.*, 1958, 42, 521-524. (b)
- LEVITT, E. E., BEISER, HELEN R., & ROBERTSON, R. E. A follow-up evaluation of cases treated at a community child guidance clinic. *Amer. J. Orthopsychiat.*, 1959, 29, 337-347.
- MAAS, H. S., KAHN, A. J., STEIN, H. D., & SUMNER, DOROTHY. Socio-cultural factors in psychiatric clinic services for children. *Smith Coll. Stud. soc. Wk.*, 1955, 25, 1-90.

(Received April 13, 1959)

DEVIANT RESPONSES IN COLLEGE ADJUSTMENT CLIENTS: A TEST OF BERG'S DEVIATION HYPOTHESIS¹

AUSTIN E. GRIGG AND JOSEPH S. THORPE
University of Texas

Berg, in a series of provocative papers (1955, 1957, 1958), has insisted that individuals when taking almost any type of non-achievement test make many responses which are similar to responses made by most others, but they also make responses which are unlike those of others in the general population and are more like responses made by members of some special subpopulation to which the particular individual belongs. Thus, some individuals respond to test items of various kinds in a manner more typical of lawyers than of engineers or of the general population. Some make responses more typical of psychotics than of psychopaths or of the general population.

Berg believes that the particular test content is not important, but that the use of clearly delineated criterion groups and an analysis of response tendencies peculiar to the particular group are the basic essentials in using deviant responses to study personality. He hypothesizes that the tendency to give deviant responses is a generalized one and is exhibited in situations other than the testing situation. In regard to personality maladjustment, he postulates, in effect, that those who give maladaptive responses in everyday living will give "deviant" responses when scored against a key of typical responses for most any kind of psychological test.

Using Berg's approach, Barnes (1956) found that a simple count of the total number of atypical responses on the MMPI correlated .87 with Scale 7 and .93 with Scale 8. In an earlier study, this same author (1955) was able to discriminate between a large

sample of controls and psychiatric patients on the basis of how the subjects responded when checking personal preferences for a small number of abstract geometrical designs. Barnes also studied clinical subgroups by the deviant response method.

Although Berg and his students are currently championing the deviant response approach, it should be remembered that as early as 1910, Kent and Rosanoff used a "deviation" approach when they altered the Jungian word association technique and appraised results in terms of common and uncommon associations to stimulus words as diagnostic of personality maladjustment.

Following Berg's reasoning, if one takes a normal sample of college students and studies the common and uncommon responses they make when asked to describe themselves on an adjective check list and then constructs an adjective list from the most commonly and the most rarely selected words, one would predict that with a new population those who become therapy clients would check more of the rare words and fewer of the common words than would the nonclient members of the population. The present study tests the validity of this prediction.

METHOD

In February 1958, during their regular freshmen testing sessions, the entering freshmen checked their self-descriptions on the 300-item Gough Adjective Check List (1955). An item analysis was made of the responses by the 181 males and 92 females to ascertain the most commonly checked adjectives and the least commonly selected adjectives. The original analysis was made separately for each sex, but it was observed that certain "sex-biased" adjectives could be eliminated from the data and the results would be applicable to both sexes. (The word "sweet," e.g., was commonly selected by girls but

¹ This research was supported by funds awarded to the University of Texas Testing and Counseling Center by the Hogg Foundation for Mental Health.

rarely by boys.) Seventy-two items of commonly checked and uncommonly checked adjectives were selected for the revised form of the test by including all adjectives selected by 84% or more of the freshmen and also all adjectives selected by fewer than 16% of the group. The final selection of 72 adjectives consisted of 33 commonly selected words and 39 uncommonly selected words.

In September 1958, this revised 72-item adjective check list was given to approximately 1,400 entering freshmen. The freshmen were asked to check the adjectives that best described them. At the close of the academic year this pool of data was divided into those freshmen who had reported during the year for psychiatric help at the Student Health Center or to the University of Texas Testing and Counseling Center for personal adjustment counseling, those who had reported for other types of counseling such as vocational-educational counseling, and into those who did not seek any type of professional counseling or psychiatric aid. In the present study, the counseling cases were classified as to type after each interview by the counselor, and those who remained as educational-vocational did not receive assignment to personality maladjustment by their counselors. A study was made of the "deviant" responses of the psychiatric treatment group, the two counseling groups, and a randomly selected control group from the nonclient group. A deviant response was defined as each time a typical adjective was not checked (from the 33 commonly selected adjectives) or each time an atypical adjective was checked (from the 39 uncommonly selected adjectives). The four groups studied were: 24 psychiatric treatment cases, 37 personal counseling cases, 186 educational-vocational cases, and 150 randomly selected nonclient controls.

RESULTS AND DISCUSSION

The means and standard deviations of the deviant response scores for each of the four groups are presented in Table 1. The psychiatric and personal counseling groups gave more deviant responses than did the educational-vocational counseling group or the control group.

The summary of an analysis of variance of

TABLE 1

MEANS AND STANDARD DEVIATIONS OF DEVIANT RESPONSES OF PSYCHIATRIC, PERSONAL COUNSELING, VOCATIONAL COUNSELING, AND CONTROL GROUPS

Group	N	Mean	SD
Psychiatric	24	14.29	8.13
Personal	37	12.08	6.99
Vocational	186	9.98	6.67
Control	150	9.77	6.87

TABLE 2

ANALYSIS OF VARIANCE OF DEVIANT RESPONSE SCORES OF PSYCHIATRIC, PERSONAL COUNSELING, VOCATIONAL COUNSELING, AND NONCLIENT CONTROL GROUPS

Source of Variation	df	Mean Square	F
Between groups	3	186.57	3.91*
Psychiatric & Personal vs.			
Vocational & Control	(1)	(484.66)	10.15*
Psychiatric vs. Personal	(1)	(71.14)	1.49
Vocational vs. Control	(1)	(4.92)	—
Within groups	393	47.72	
Total	396		

* Significant at the .01 level of confidence.

the deviant responses is given in Table 2. The between-group means *F* test, which yielded an *F* value of 3.91, significant at the .01 level, was followed by three orthogonal comparisons between the various groups. The first of these comparisons indicates that the psychiatric group and the personal counseling group, when combined, differ significantly from the vocational group and the control group taken together, when compared for average number of deviant responses. The two groups with personality problems give significantly more deviant responses, a finding consistent with Berg's hypothesis. The remaining comparisons show that the psychiatric group does not differ significantly from the personal counseling group in deviant response scores and that the vocational group does not differ significantly from the control group in average number of deviant responses.

The differences in the data are consistent with Berg's Deviation Hypothesis. The two groups who sought professional help for personal adjustment problems, the psychiatric group and the personal counseling group, presumably exhibited more personality disturbance than did members of the nonclient control group or members of the educational-vocational counseling group. In the present study, although most of the deviant adjectives are of the type that would be selected by subjects with a response set to describe themselves in self-deprecating terms, only 39% of the adjectives in the present study occur in

Gough's list of "least favorable" adjectives, and only one of the 30 items of Gough's Adjustment Scale from his Adjective Check List (1955) occurs in the total pool of commonly and uncommonly selected adjectives of the present study.

We call attention to a novel element in our study. The traditional manner of deriving an empirical discrimination between two groups has been to give a test to two criterion groups, e.g., well-adjusted and maladjusted, select the discriminating items by an item analysis, and then to cross-validate the empirical key with new groups. In the present study, modeled to test Berg's hypothesis, the empirical test key was established by studying a general population in terms of typical and atypical responses. Then various criterion groups drawn from a new population were used to ascertain the validity of the deviant response key.

CONCLUSION

A list of commonly selected and uncommonly selected adjectives from the Gough Adjective Check List was established on the basis of responses of a group of university freshmen. When this list of frequently chosen and infrequently chosen words was given to

a new sample of freshmen, it was found that the deviant response scores of freshmen who later became personal counseling or psychiatric treatment cases were significantly greater than the deviant response scores of a control group of nonclients. A group of freshmen who later applied for educational-vocational counseling and were retained in that classification by their counselors did not deviate significantly from the control group in their deviant response scores. These results are consistent with Berg's Deviation Hypothesis.

REFERENCES

- BARNES, E. H. The relationship of biased test responses to psychopathology. *J. abnorm. soc. Psychol.*, 1955, **51**, 286-290.
- BARNES, E. H. Factors, response bias and the MMPI. *J. consult. Psychol.*, 1956, **20**, 419-421.
- BERG, I. A. Response bias and personality: The Deviation Hypothesis. *J. Psychol.*, 1955, **40**, 60-71.
- BERG, I. A. Deviant responses and deviant people: The formulation of the Deviation Hypothesis. *J. counsel. Psychol.*, 1957, **4**, 154-161.
- BERG, I. A. The unimportance of test item content. In B. M. Bass & I. A. Berg (Eds.), *Objective approaches to personality assessment*. New York: Van Nostrand, 1958.
- GOUGH, H. G. *Reference handbook for the Gough Adjective Check List*. Berkeley, Calif.: Univer. Calif. Inst. Pers. Assessment & Res., 1955.

(Early publication received July 15, 1959)

BRIEF REPORTS

INDIVIDUAL AND SOCIAL CORRELATES OF PRISON ESCAPES¹

D. A. DOBBINS,² F. E. STOCKWELL

Louisiana Department of Institutions

AND W. S. LOVING

Louisiana State Penitentiary

Many prison administrators believe prison escapes are primarily situational, based on opportunity and triggered by stresses of the prison environment. The fact remains that some inmates will escape and others will not even when opportunities are equal. This study attempts to identify some personal and social characteristics of the "escape risk" inmate and, in so doing, help fill a curious scarcity of empirical research on the subject.

The Ss were 200 white male inmates of the Louisiana State Penitentiary. Half had successfully escaped over a two-year period. The remaining 100, matched by month and year of admission and degree of custody, were nonescapees. Each escapee was thus matched with a non-escapee imprisoned for the same length of time and under the same degree of supervision at the time of escape.

Twenty-two variables were subjected to chi square analysis. Eight variables discriminated the criterion groups above the .05 confidence level. Traits representative of the escape groups were as follows: fewer dependents, more prior out of state penal commitments, more property crimes committed, more juvenile commitments, younger

at first arrest, from smaller cities, from other states, and fewer years residence within Louisiana.

Three additional variables were significant between the .05 and .10 levels. Traits predominant in the escape group were: longer sentences, lower tested educational achievement, and younger at admission.

Variables which did not discriminate between the two groups were: Otis IQ, marital status, broken homes, urban backgrounds, father's occupation, number of siblings, previous prison commitments within state, veteran status, psychiatric history, years school completed, and number of institutionalized siblings.

Significant variables were intercorrelated and cluster analyzed using Tryon's modification of Holinger and Harman's *B* coefficient. The first cluster, with a *B* coefficient of 2.41, included the following variables: years residence within Louisiana, distance to home state, number penitentiary commitments elsewhere, and size of community. This cluster was labeled "Transient Criminality."

The remaining four variables were included within one additional cluster with a *B* coefficient of 2.64. The cluster, identified as "Early Criminal History," included number of juvenile commitments, age at first arrest, number dependents, and commission of property crimes.

The results suggest that with situational factors held reasonably constant, the risk of a given prisoner's escaping is related to certain precommitment life patterns.

(Received April 20, 1959)

¹ An extended report of this study may be obtained without charge from F. E. Stockwell, Louisiana Department of Institutions, Baton Rouge Louisiana, or for a fee from the American Documentation Institute. Order Document No. 6114, remitting \$1.25 for microfilm or \$1.25 for photocopies.

² Now with the Personnel Research Branch, TAGO, Washington, D. C.

A DIMENSIONAL ANALYSIS OF CLINICAL JUDGMENT¹

JOHN W. DONAHOE

University of Kentucky²

Commonly, studies of clinical judgment severely restrict the mode of response open to the judge (e.g., psychiatric diagnosis). This is done in order to yield a readily quantifiable dependent variable (e.g., number of correct classifications). Unfortunately, such decisions constitute an unrepresentative sample of the kinds of judgments that clinicians are called upon to make. Rather, what appears to be needed is an instrument allowing for more diversified responses so as to provide a knowledge of the dimensions along which clinical judgments vary. The present study was an attempt to develop a rating scale to fulfill this function.

The rating consisted of 50 *descriptive* and 25 *dynamic* seven-step scales. The *descriptive* scales were chosen from adjectives and phrases used by Beck in the interpretation of test materials. Terms of opposed meanings (e.g., *constricted-spontaneous*) were selected as scale endpoints. The 25 *dynamic* scales dealt with the chief sources of conflict and the mechanisms of defense according to psychoanalytic theory (e.g., *never uses projection-very frequently uses projection*).

The Rorschach productions of six hospitalized mental patients were scored and summarized. The diagnostic categories represented were schizophrenia, depressive reaction, personality trait disturbance, and chronic brain syndrome. Twelve

clinical psychologists with an average of 3.1 years of diagnostic experience served as judges. All judges were successively presented with the six protocols and asked to place a check mark on each of the 75 scales and also to diagnose the patient in accord with conventional psychiatric nomenclature.

The median interjudge reliability coefficient was .51, with a range from .18 to .98. In order to examine the structure of the rating form, a scale matrix was constructed. Those scales with squared vector lengths of less than 5.0 were not included, thereby reducing the number of *descriptive* scales from 50 to 42 and the number of *dynamic* scales from 25 to 6. The reduced matrix was analyzed by the *D* method of factoring. The first factor was defined by the scale endpoints *much hostile conflict* (pivot), *neurotic*, *unstable*, *insecure*, and *anxious*. These are terms intimately associated with the concept of neuroticism. The second factor dealt with a continuum from introversion to extroversion as the scales having highest coordinates were *cautious* (pivot), *introversive*, *constricted*, and *withdrawn*. The third factor, while not as homogeneous as the others, was concerned with psychosis. Scales heavily loaded on the third factor were *organic brain involvement* (pivot), *defenses holding poorly*, *psychotic*, and *underreactive*.

The judges were correct in only 38% of their diagnoses. When the matrix was analyzed with respect to Rorschach protocols, however, the records were appropriately grouped. Thus, while the judges were capable of checking the rating form with some validity, they were unable to integrate these separate judgments into an accurate global statement. In short, the scaling technique should provide a means for assessing clinical judgment under a wide variety of conditions.

(Received June 12, 1959)

¹ An extended report of this study may be obtained without charge from John W. Donahoe, Psychology Department, University of Kentucky, Lexington, Kentucky, or for a fee from the American Documentation Institute. Order Document No. 6115, remitting \$1.25 for microfilm or \$1.25 for photocopies.

² This study was conducted while the author was employed at the Veterans Administration Hospital, Lexington, Kentucky.

SOME CORRELATES OF MANIFEST ANXIETY IN CHILDREN¹

IRA ISCOE AND IRENE COCHRAN

University of Texas

Two hypotheses were tested. 1. If a relationship exists between anxiety as measured by the CMAS and observable behavior, then children who score high on the CMAS should be judged as significantly more "maladjusted" than their classmates of the same sex who score low on this scale. 2. If Manifest Anxiety is related to adjustment, it should bear a significant relationship to the California Test of Personality, a frequently used instrument to assess "adjustment" in school age children.

For Hypothesis 1, 118 boys and 96 girls, average age 10.4 years, from six elementary school classrooms, were administered the CMAS. For each grade the five boys and five girls having the highest and lowest anxiety scores were selected. A total of 60 Ss was thus obtained. Mean HA was 25.5 and LA was 3.4. Homeroom teachers rated each S on "adjustment" compared to own sex group, using a five-point rating scale. They were unaware of CMAS scores of the students. χ^2 was equal to 8.2 ($df = 2$) which is significant at the .02 level. A rank order correlation yielded a rho of .37, significant at .01 level. Manifest Anxiety would appear to be an observable characteristic, noticeable to those familiar with the Ss' everyday school behavior. Since the concept of Manifest Anxiety is derived from Hullian theory, a consideration of the

relationship between high and low drive and teachers' judgments seems appropriate. Questioning of teachers as to the basis of their judgments revealed a tendency to equate "maladjustment" with "restlessness," lack of attention, and inability or unwillingness to "settle down." Good adjustment indicated the opposite tendencies. There is some support, therefore, for the belief that drive level was a prominent observable characteristic.

For Hypothesis 2, 228 fourth- and sixth-grade children were administered the California Test of Personality and the CMAS. High scores on the former indicate good adjustment.

CMAS correlated $-.48$ with "Total Adjustment," $-.49$ with "Personal Adjustment," and $-.59$ with "Freedom from Nervous Symptoms." These r 's are highly significant ($p = .01$).

The results of Hypothesis 1 suggest a relationship between general body activity and judgments of adjustment. This is commensurate with the concept of Manifest Anxiety as a measure of drive level. The results of Hypothesis 2 suggest a significant relationship between high drive level and adjustment and also the willingness to admit certain subjectively experienced discomforts and apprehensions. The possibility of similarity of item content on both instruments should also be considered.

Despite certain shortcomings, the concept of Manifest Anxiety in children as a measure of drive level, if not of adjustment, has in it potentialities that deserve further study.

(Received August 9, 1959)

¹ An extended report of this study may be obtained without charge from Ira Iscoe, Department of Psychology, University of Texas, Austin 12, Texas, or for a fee from the American Documentation Institute. Order Document No. 6118, remitting \$1.25 for microfilm or \$1.25 for photocopies.

A CLUSTER ANALYSIS OF OBJECT SORTING BEHAVIOR¹A. B. SILVERSTEIN²*Pacific State Hospital*

Clinical psychologists are frequently called upon to assess a patient's ability to think abstractly, and we possess a number of instruments with which we endeavor to measure this ability and gauge its impairment. Not much is known, however, concerning the relationships among the measures which these instruments yield. The aim of the present study was to clarify the interrelations of certain scorable aspects of the Goldstein-Scheerer object sorting test as used by Rapaport, Gill, and Schafer (1945).

The data were those reported by Rapaport et al. for 54 randomly selected members of the Kansas Highway Patrol.³ Thirteen variables were selected for analysis. All distributions were dichotomized as close to the median as possible, and ϕ/ϕ_{\max} was employed as the index of association. McQuitty's elementary linkage analysis was then utilized to treat the data.

The analysis yielded three clusters, two of which correspond closely to Part I (active concept formation) and Part II (passive concept formation) of the object sorting test. One exception was that Concrete Definitions from Part II of the test appeared in the cluster with the active concept-formation variables. The second

exception was that Functional Definitions from both parts of the test formed a cluster of their own.

Rapaport et al. consistently distinguish between active and passive concept formation and place some emphasis on the diagnostic significance of differences in performance on Parts I and II of the test. The findings of this study indicate that the distinction between the two parts of the test is quite justified. Not all research workers make this distinction, however. A comparison of performance on Parts I and II as scored by the "conceptual area" analysis of McGaughran and Moran (1956) would appear highly desirable in the light of the present findings.

It has been suggested that more valid estimates of intelligence, e.g., on Wechsler's vocabulary, might be obtained if functional definitions were to receive less credit than concrete definitions since the order abstract-concrete-functional appears to be closer to the truth than the conventional order abstract-functional-concrete. In contrast to this, the results of the present analysis suggest that while abstract and concrete definitions may lie on a continuum, functional definitions actually represent a thing apart.

REFERENCES

- McGAUGHRAN, L. S., & MORAN, L. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and nonpsychiatric groups. *J. abnorm. soc. Psychol.*, 1956, 52, 43-50.
- RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. 1. Chicago: Year Book, 1945.

(Received August 20, 1959)

¹ An extended report of this study may be obtained without charge from Arthur B. Silverstein, Pacific State Hospital, Pomona, California, or for a fee from the American Documentation Institute. Order Document No. 6117, remitting \$1.25 for microfilm or \$1.25 for photocopies.

² Formerly at the Psychiatric Institute, University of Maryland.

³ The author is indebted to D. Rapaport and to the Year Book Publishers for permission to make use of material appearing in *Diagnostic Psychological Testing*.

PSYCHOLOGICAL TEST REVIEWS

New Tests

Stott, D. H., and Sykes, Miss E. G. *Bristol Social Adjustment Guides*. Age 5-15.¹ Adjustment guide (observation blank), scoring templates, and diagnostic form (summary record blank) for (a) the child in school (boy) BG-1, and (girl) BG-2; (b) the child in residential care BG-3; and (c) the child in the family BG-4. Guides BG-1 and BG-2 use the same scoring templates and diagnostic form. Guide BG-4 includes a life history blank and is scored using a copy of the guide with overprinted key in lieu of templates.

The *Bristol Social Adjustment Guides* constitute a careful attempt to render more objective and precise the observations of teachers, social workers, and foster parents, with respect to children's social adjustment. The *Guides* are intended to save time for the psychologist by providing accounts of behavior more standardized and compact than anecdotal protocols, and more specific than impressionistic ratings. Since the *Guides* encourage accuracy in observing children and provide a system for classifying symptomatic behavior, they have a place also as aids in clinical training.

Each *Guide* consists of numerous adjectives and phrases grouped by topic and scrambled as to desirability, though with negative indications in the majority. The observer simply underlines those terms which apply to the child concerned. Translucent templates are provided on which critical items are keyed for maladjustment in specific areas such as "hostility to adults," "anxiety," "knavery," and "psychosomatic symptoms." The results can be quasi-quantified on the "diagnostic form" on which the keyed items are recorded, by counting the significant items marked in each category. Items in certain areas are also grouped as to severity. However, the author eschews any overall index of maladjustment.

Stott and Sykes are thoroughgoing empiricists, and their work is a good example of an inductive approach to conceptualizing maladjustment. Avoiding the constructs of current theories, they collected descriptions of specific children as reported in case conferences. By successive tryouts with teachers, social workers, etc., they developed lists of descriptive terms which could be used comfortably and unambiguously by nonpsychologists. Items were grouped into approximately 15 "attitudes," e.g., "depression" or "hostility to adults," by a scalogram-like procedure in which children's names on one axis of a large grid and their symptoms on the

other were ordered and recorded until coherent syndromes of behavior emerged. Only those items are now scored whose relative frequencies among the stable vs. the maladjusted (omitting a middle category of "unsettled children") yielded significant χ^2 's. However, the "attitude" syndromes shade into one another along a behavioral spectrum and have merely descriptive significance.

Since the *Adjustment Guides* are intended only to elicit and organize the specifics of observed behavior, not all the usual criteria of test evaluation are relevant. The 46-page manual by Stott includes a lengthy defense of his scientific philosophy and a detailed historical account of the successive steps in the development of the *Guides*. Their reliability (interrater agreement) in a preliminary study is given as around .76. More data are needed, but reliability is likely to depend as much upon the raters' skills and opportunities as upon the content of these forms. The validity of each item in differentiating between the stable and the maladjusted is presented, but without cross-validation. Normative data for the interpretation of total scores on each attitude are lacking. This is not a serious omission for American users who will need to develop their own norms anyway.

Differences between the English and American languages constitute a more fundamental problem. The author's success in finding meaningful descriptive phrases drawn from the colloquial speech of nurses, teachers, etc., e.g., "spivvish dress," or "feckless," will occasionally create problems of understanding for Americans. Even clinical vocabulary has its differences—e.g., "unforthcomingness" for "shy," or "unsettled" for "mildly maladjusted." Those of us used to the euphemistic polysyllables of "psychopathic deviate" may feel decidedly self-conscious characterizing a child as a "knave." The *Guides*, in summary, will probably not achieve their potential value in this country until someone undertakes to translate them from the authors' English into ours.—Read D. Tuddenham.

Carrol, J. B., & Sapon, S. M. *Modern Language Aptitude Test*. Ages 14—adult. 2 forms, long (60 min.) and short (30 min.). Hand or IBM scoring. Tape recording necessary for administration of long form (\$7.50); test booklets (\$3.50 for 25), IBM answer sheets (\$3.50 for 50), scoring stencils (\$0.50 per set), specimen set (without tape) (\$0.75), manual, 27 pp. New York: Psychological Corp., 1959.

The *Modern Language Aptitude Test*, the outcome of a five year study, includes aural and visual ma-

¹ With manual, *The Social Adjustment of Children*, pp. 46. London: University of London Press, 1958.

terials. The long form which contains both types of materials requires a tape recorded for administration. The two aural parts of the test, included only in the long form, are "Number Learning" and "Phonetic Script." The other three are: "Spelling clues," requiring recognition of incompletely spelled words; "Words in Sentences," testing sensitivity to grammatical structure; and "Paired Associates," intending to tap memory.

Sufficient data bearing on reliability and validity are presented to warrant issuing the test and to establish confidence in its usefulness as a diagnostic device. The normative data, particularly for high school and adult norms, are weak.—E. S. B.

Stark, R. H. *Polyfactorial Study of Personality*. Adults. Test booklets (\$6.25 per package), IBM answer sheets (\$2.75 per package), stencils (\$2.00), profile sheets (\$2.75) specimens set (\$2.50), manual, 11 pp. (\$0.75). New Rochelle, N. Y.: Bruce, 1959.

A 300-item inventory reportedly selected on the basis of factor analysis and correlations with unspecified criteria. The general impression is that the major criteria may have been other tests, e.g., Rorschach, MMPI, and the H-T-P test. Ten scores are obtained and Tables of Correlations with MMPI, the Association Adjustment Inventory, and certain subtests of the WAIS are presented. Test-retest reliability for a prison population suggests that the scores are reasonably stable. No other validation data are offered, even though reference is made to various possible criterion populations. There are no references to published reports. Why was this test issued after so little developmental analysis?—E. S. B.

Cassell, R. N. *Test of Social Insight*. Two forms, youth edition (junior and senior high school) and adult edition. Untimed. Test booklet (\$6.25 per package); IBM answer sheets (\$2.75 per package); specimen sets (\$2.50); Profile sheet (\$2.75 per package); IBM scoring stencils (\$1.00), manual, 30 pp. (\$1.75). New Rochelle, N. Y.: Bruce, 1959.

The *Test of Social Insight* consists of 60 five alternative multiple choice items dealing with home and family relations, authority figures and social agencies, play and avocational interests, and work and vocational interests. The two editions are parallel in content, only slightly altered to adapt to the different age groups. In addition to a total score, five part scores: withdrawal, passivity, cooperativeness, competitiveness, and aggressiveness,

are obtained. There are two sets of norms for the youth edition (junior and senior high school students) and two sets for the adult edition (college students and adults and guidance counselors and personnel psychologists).

The odd-even reliabilities, even after Spearman-Brown corrections, are unsatisfactory, none above .88 and, in some populations, as low as .51. Validities based on inadequately reported unpublished data are more reassuring. Caveat Emptor!—E. S. B.

Bruce, M. M. *Bruce Vocabulary Inventory*. Adult. Untimed (20 min.); Specimen sets (\$1.00); Test booklet (\$5.00 per package); IBM answer sheet (\$2.75 per package); manual, 3 pp. (\$0.75); scoring keys, hand sets (\$2.00), and machine sets (\$4.50). New Rochelle, N. Y.: Bruce, 1959.

A 100-item multiple choice test in which the key words were chosen at random from the even numbered pages of the *Oxford Universal Dictionary*. Norms are given for eight not very precisely defined occupational groups and for the aggregate. The groups vary from 61 to 495 in size. While the reported reliabilities, odd-even, .92, and test-retest, .84, are acceptable, the test correlates only slightly more (approximately .80) with other vocabulary tests than with intelligence tests (approximately .60).—E. S. B.

Bruce, M. M. *Association Adjustment Inventory*. Adult. Untimed (10 min.); Test booklet (\$5.00 per package); IBM answer sheets (\$2.75 per package); scoring keys, hand set (\$2.00) and machine set (\$4.50); manual, 15 pp. (\$0.75). New Rochelle, N. Y.: Bruce, 1959.

The Kent-Rosanoff list of stimulus words has been cast up into a four alternative multiple choice format. Thirteen different scores are obtained, including a total (maladjustment) and an immaturity score, and 11 dimensions of "deviate ideation." Tables of intercorrelation raise serious doubts as to whether there are that many independent scores. Norms for "institutionalized" and "general" populations are based on large but vaguely defined populations. Test-retest correlations for one week and one month intervals are satisfactory, but the base for claiming validity is very unsatisfactory. There is evidence of some general relationships with MMPI scales and with a largely unvalidated inventory. Reference is made to unreported results of diagnostic validity studies that are unpublished. It should have been marked "For Experimental Use."—E. S. B.

INTELLIGENCE: A CONCEPT IN NEED OF RE-EXAMINATION¹

SHEPHARD LIVERANT

Ohio State University

The importance of the concept of intelligence in understanding and predicting human behavior appears so obvious that the value of this term in a science of behavior has seldom been questioned by psychologists. The inadequacies inherent in present conceptions of intelligence have been made explicit by a number of authors (Anastasi, 1958a; Coombs, 1952; Fromm & Hartman, 1955; McClelland, 1951), but the reader is left with the impression that the concept itself is essentially a useful and necessary one. By leaving open the possibility that the concept may be unnecessary, and even an obstacle, in explaining behavior, perhaps we can arrive at a better solution to the problems now subsumed under the rubric of intelligence.²

This paper explores a number of problems over which the intelligence concept has stumbled, such as: establishing an unequivocal definition, the contribution of heredity and environment, measuring an innate capacity, situational predictions, the constancy of the IQ, intelligence as a general factor, the role

¹ This paper was written in part at the Interdisciplinary Program in the Behavioral Sciences at the University of New Mexico, Summer 1958. The conference was sponsored by the Behavioral Sciences Division, Air Force Office of Scientific Research under Contract AF 49 (638)-33. Their support is gratefully acknowledged.

The author also wishes to express his appreciation to Julian B. Rotter, Richard Jessor, and Edwin Barker for their invaluable suggestions and criticisms.

² In order to forestall misunderstanding, it is emphasized at the outset that the term intelligence is being questioned in its explanatory usage, not necessarily as a descriptive concept. In other words, intelligence is used in two ways; as a descriptive term (e.g., A's behavior is more intelligent than B's) and as an explanation (e.g., A solved more problems than B because A is more intelligent). It is the validity of the latter use which is being questioned.

of culture in defining intelligent behavior, and the function of nonintellective variables. The objective of such an exploration is to re-evaluate these persistent problems by analyzing the logical and methodological adequacy of the meanings carried by intelligence.

Present uses of the term intelligence for individual prediction and the difficulties inherent in such usage are best represented in the clinical setting. Consequently a brief look at the clinical approach to intelligence will serve as a point of departure for the subsequent analysis.

CLINICAL APPROACH

Clinical case studies have produced a greatly increased awareness of the failure of an IQ (or any other behavioral appraisal) to reflect accurately a constant ability to perform at a certain level in all situations. This awareness has led to the clinical practice of considering the IQ as an index of present intellectual functioning rather than intellectual potential or capacity. However, the belief in the existence of a "true" IQ or intelligence persists, and attempts to discover it in each individual continue.

Largely on the basis of scatter or pattern analysis of intelligence tests, statements are made concerning the size of the discrepancy between what is now present in the way of intellectual functioning and some assumed absolute capacity. Were uneven test profiles produced only by deviant individuals such interpretations might be meaningful, but subtest scatter is the rule rather than the exception within any population in spite of efforts to build homogeneous measures (Shofield, 1952). As will be argued later, it is probably more fruitful and certainly more empiri-

cally justifiable to regard uneven test profiles as the result of conditions (not necessarily abnormal ones) which have differentially affected the acquisition of certain skills and/or the manifestation of these skills in certain classes of situations. Then the task becomes one of determining what these conditions are and how they operate, rather than ascribing subtest patterns to various types of disease entities which affect another entity (intelligence).

In general it appears that the host of evidence gathered over the last twenty or more years challenging the general and innate nature of intelligence has been subsumed under an interference hypothesis. Consequently, intelligence can be and still is viewed as a general and innate source of variance, but it is increasingly recognized that in order to get at it from behavioral observation one must strip away an ever increasing number of complex variables which interfere with its effects on behavior.

This point of view is illustrated by the work of the Chicago group (Eells, Davis, Havighurst, Herrick, & Tyler, 1951) who, after demonstrating the effects of culture on test performance, conclude that the answer to the problem is to develop "culture-free" intelligence tests. By contrast, Anastasi (1950) uses the same kind of evidence to question the validity of any test as a measure of an abstract conception of intelligence without regard to validation criteria. As Rotter (1958, p. 19) points out, "It is somewhat surprising that the emphasis on culture and training did not result in a discarding of the notion of over-all intelligence and the IQ in favor of a clear understanding of the value of varied tests in assessing what has been learned and what each test can predict. Rather the early trend to entitize and reify the IQ continues." (Quoted by permission of McGraw-Hill.)

The question being raised here is what if anything will be left as a contributor to understanding behavior on a psychological level once the stripping away process is complete. If more and more of the variance which accounts for individual differences in intellectual performance can be attributed to other variables, what is intelligence?

THE PROBLEM OF DEFINITION

Spiker and McCandless (1954) attempt to resolve the definitional problem by an appeal to early operationalism. In effect they conclude that intelligence is what different intelligence tests measure in different situations with different groups and the relationship of these measures to other measures. Their solution reduces intelligence to the status of an intervening variable or pure disposition term as described by Carnap (1956). There is nothing wrong with such a definition of intelligence except that most people are not willing to accept it. Intelligence means more to most of us than a score on an intelligence test even though the "more" hasn't been made explicit (i.e., we use intelligence as a hypothetical construct whose validity depends on the verification or refutation of predictions derived from the theory which contains it).

More recent thinking on the subject suggests that the ultraempirical approach is not a satisfactory solution to the problem of definition. Several authors (Cronbach & Meehl, 1955; Jessor & Hammond, 1957; Liverant, 1958) have argued that the construction, interpretation and validity criteria of any test depend on the nomological network (meanings) surrounding the implicit or explicit construct(s) which the test is designed to measure. Intelligence like any other construct takes its meaning from the assumed antecedent conditions and subsequent responses which it mediates. The controversies over nature-nurture on the antecedent side and general versus specific factors on the response side of the relationship characterize the difficulties in arriving at an accepted definition of intelligence. Thus, the next step is to examine first the nature of the antecedent conditions and then the subsequent responses subsumed by intelligence.

ANTECEDENT SIDE OF THE PROBLEM

Heredity and Environment

As now evolved, the antecedents of intelligence do not refer to a set of conditions (either inherited or environmental) but rather to a conceptual model involving postulated genetic operations. Although stated in many different forms, genetic intelligence seems to

embody the idea that each organism, from the moment of conception throughout the life span, has inherent within it an unmodifiable growth process determined by gene composition which if given the proper environmental stimulation at the right time will result in maximum problem solving performance for that organism under suitable conditions of motivation, health, etc. Improper stimulation (degree and kind unspecified) may inhibit certain manifestations of this innate property, but given an adequate restoration of environmental conditions (provided structural damage has not occurred) it will proceed upon its genetically determined course. However, no amount of environmental manipulation will serve to modify the limits of intellectual performance prescribed by the genetic constituents. It is this last proposition which lends an air of crucial significance to this formulation, since knowledge of these innately fixed limits (intelligence) will permit us to predict maximum attainment under optimal conditions—predictions which are invaluable in terms of educational and vocational planning.

There are several implications of genetic intelligence which need clarification, if not complete refutation, in light of recent conceptual advances in both genetics and psychology. Paramount is the notion of intellectual limits set by heredity. As usually interpreted this idea means that beyond a certain point (usually unspecified) any and all environmental conditions cease to make a difference as far as increments in problem solving proficiency are concerned. But the conclusions of modern day geneticists (e.g., Huxley, 1949; Stern, 1952) reveal a basic fallacy inherent in this line of reasoning. Geneticists postulate that the appearance of any characteristic (behavioral or otherwise) depends upon the *interaction* of a *specific* genetic structure operating within a *specific* environment. In this sense genes set the limits within any given environment but these limits may vary as the environment varies. Without in any way arguing against the Wiesman-Morgan theory of the immutability of the germ plasm, it is no more logical to attribute a limiting role to heredity than it is to assign the same role to the environment. Perhaps the irrevocability of genetic determi-

nants as compared to situational ones has contributed to this mode of thought, but it is one which contradicts the postulated interaction effect.

The implication of a true interaction is not that only within a limited range (set by one of the variables) does the effect of the other make a difference, but that they both operate to produce their effects over any conceivable range. Now it does appear that heredity can influence the variety of conditions which will or will not have an appreciable effect on a given characteristic. For example, it can be said that species characteristics are determined by heredity in the sense that the occurrence or nonoccurrence of an extremely wide range of external conditions does not make a difference in the morphology necessary to identify a given species. Both theoretically and empirically, however, a set of conditions could be and, in fact, have been produced which radically alter species characteristics. Feeding thyroid extract to the young axolotl to produce a creature adapted for the land rather than the usual aquatic life is a case in point (Jennings, 1930). On the other hand, intraspecies behavior is apparently affected by a much greater host of non-genetic variables. Furthermore, the importance of situational variations in determining individual differences seems to increase as one ascends the phylogenetic scale. Beach's (1954) findings regarding comparative variations in sexual behavior emphasize the ever increasing role of situational factors in determining the sexual patterns displayed by increasingly complex organisms. It is likely that the same relationship is even more predominant in the case of complex problem solving behavior. Since the evidence points to such a great number of circumstances which can affect intellectual performance it becomes conceptually difficult to speak of heredity as the determiner of intellectual limits, particularly in the case of human beings.

From this line of reasoning it follows that although no practical way may be known to affect appreciable positive changes in an individual's performance level, it cannot be assumed that such changes are precluded by genetically determined limitations. On the contrary, the implication of genetic theoriz-

ing in this respect is that changes in any characteristic would result from sufficient environmental changes. After producing previously unknown patterns of performance in pigeons (i.e., bowling behavior), Skinner (1958, p. 96) concludes that, "It is dangerous to assert that an organism of a given species or age *cannot* solve a given problem." The implications of this reasoning in the field of mental deficiency are apparent.

A further failure to recognize the significance of the continuous interaction between genes and environment in producing behavior is indicated by the one-way effect of interaction implied by the intelligence model. The implication is that training can only have an effect after neural maturation paves the way and that it is by controlling the maturational rate that heredity plays its vital role in the development of intelligent behavior. The "Johnny and Jimmy" type of studies (McGraw, 1935) are cited as evidence in this connection. But this approach neglects the reciprocal relationship between physiological maturation and behavior. The "learning how to learn" phenomena demonstrated by Harlow (1949) strongly suggests that the rate of acquisition of problem solving ability is directly dependent on neural changes which occur as a function of previous problem solving performance. In other words, previous experience as well as genes "shapes up" the organism's maturational readiness for new training procedures.

Acceptance of the interaction thesis concerning heredity and environment invalidates the concept of innate capacity unless the term includes the operations of an environment. Consequently, the prediction of behavior based on genetic antecedents involves the problem of specifying situational conditions, a problem with which the present definition of intelligence as an inherited capacity has not dealt. There are two kinds of situational conditions which need to be specified: first, the conditions of development, growth and learning, and second, the criterion conditions (i.e., the actual performance situations). The emphasis in this section is on the former conditions. The criterion conditions will be discussed more fully in the subsequent response section.

Environmental Specificity

Perhaps the most pronounced conceptual weakness of the present model lies in its inability to cope with environmental specificity. The position that individual differences in problem solving are a function of heredity given a relatively constant "demographic" environment (e.g., similar cultural opportunities or similar family life) grossly oversimplifies the specific and still unknown effects of the "psychological" environment on the hereditary potential.

Recent research concerned with the effects of early experience upon later behavior has brought into sharp focus the preeminent role of experiential factors in influencing problem solving processes. For example, evidence has been gathered that rats (Beach & Jaynes, 1954), dogs (Clarke, Heron, Fetherstonhaugh, Forgays, & Hebb, 1951), and humans (Stone, 1954) reared in impoverished perceptual and social environments show "irreversible" problem solving impairment in adulthood. Extensive clinical case material now supported by still tentative research findings (Axline, 1949) indicates that much more is involved in a child's inability to profit from the most "well conducted" training procedures than lack of the proper genetic constituents. The crucial role of interpersonal relationships in the acquisition of problem solving behavior is becoming increasingly evident.

In order to avoid confusion, it is important to recognize that no amount of empirical evidence relating intelligent behavior or the lack of it to environmental events can directly invalidate the model in question. As now conceptualized the position concerning genetic determinants is that they exist and somehow make a difference, a position which it would be absurd to deny. The most that data can accomplish is to question the usefulness of the genetic intelligence scheme and indicate lines along which a potentially more productive one may develop. The network of studies beginning to elucidate the complex set of interrelated past and present situational factors involved in determining the adequacy of intelligent behavior tends to place a heredity explanation of problem solving ability in the position of an appeal to ignorance. In

the present stage of conceptualization in this area, it is the absence of knowledge concerning situational determinants which brings forth genetic ones.

The Constancy of the IQ

In the final analysis the value of any scientific approach must be judged in terms of the number of valid hypotheses which it generates. The questionable validity of one of the two major hypotheses stipulated by the genetic theory of intelligence (i.e., intelligence as a general factor) will be discussed in the subsequent response section. The inability to validate or refute the other major hypothesis concerning the constancy of the IQ represents still another instance of the essential untestability of the genetic model.

It may well be that given a theoretically constant environment individuals would maintain the same relative position in problem solving adequacy over any given time period as a function of genetic differences. However, the indications that the *inconstancy* as well as the *constancy* of the IQ can be better explained by alternative hypotheses (other than genetic ones) seriously challenge the existing model. Disregarding individual differences in patterns of IQ change, the empirically observed "constancy" of intellectual status at the upper childhood age ranges has been accounted for on the basis of an "overlap" hypothesis. Anderson (1940) among others has proposed that the similarity in IQ from one age level to the next depends on the proportion of similar test items included in the total score at various age levels.

A more crucial test of the constancy hypothesis is provided by the longitudinal studies (Bayley, 1955) explicitly considering individual differences. These conclusively demonstrate the lack of IQ constancy over any childhood period. Sontag, Baker, and Nelson (1958), in probably the most thorough and well designed study of its kind, found that 62% of their children changed more than 15 IQ points in either direction sometime during the developmental period from the age of three through ten years. Most significant is the finding that the direction and degree of change in rate of "mental growth" are re-

lated to personality dimensions (primarily a need for achievement).

Reductionism

It can and has been argued that environmental explanations in and of themselves must always remain incomplete since genetic differences undoubtedly do play a role. The demand for the incorporation of genetic operations in behavioral explanatory conventions brings us to the problem of reductionism, i.e., the question of the necessity and/or feasibility of seeking explanations of the phenomena described in the language of one discipline by reference to the descriptive language of another. Reductionism, as far as psychology is concerned, usually refers to attempts to explain molar behavior in terms of physiological (i.e., neurological) events. Since genetics is still lower in the hierarchy, in the sense that its effects on behavior must be mediated through physiological processes, the criticisms cited in reference to physiological reductionism should be even more pertinent in the case of genetic reductionism. Jessor (1958) cogently argues against the logical possibility of reducing constructs designed to abstract organism-environment interactions to physiological operations, since the latter lack terms to describe the behavioral environment. This argument seems to be particularly applicable to any attempts to seek the causal antecedents of problem solving behavior (certainly a complex organism-environment interaction) in genetic laws (certainly lacking terms to describe the behavioral environment).

Whatever the eventual relationship between genetic determinants and complex behavior proves to be, it is obvious that this relationship is a very distant one.³ Bridging this gap necessitates taking into account a number of intervening processes at different levels of organization. And the simpler or more molecular the process, the closer is the relationship to the genes. Thus, initial sensitivity to stimuli may be more genetically determined than differential responding to these stimuli. Since these processes preceding complex performance are continuously changing as a function

³ Anastasi (1958b) elaborates this same point in much more detailed and analytic fashion in a recent article.

of innumerable environmental variations in large part brought about by the behaving organism, the more complicated the behavior, the greater is the potential error in prediction based on genetic antecedents.

In this sense, it is not at all surprising to find a lack of correlation (Bayley, 1955) between infant scales which of necessity sample lower (more molecular) levels of behavior and intelligence tests dependent on higher order processes. But the major conclusion to be drawn from our attempts to develop infant scales and culture free tests is that the more we narrow our performance sample to eliminate or minimize the effects of environmental variables, the less likely we are to arrive at valid measures of intelligent behavior as we now conceptualize it.

The enormous difficulties involved in establishing predictive relationships between genes and complex behavior raise the question of the usefulness of any single construct which attempts to mediate such a relationship. Fuller's (1954, p. 16) statement is relevant in this connection, "On a *priori* grounds a biologist would suspect that intelligence would not be a very good subject for genetic analysis, since it is a complex character depending upon the coordination of many separate processes."

SUBSEQUENT RESPONSE SIDE OF THE PROBLEM

Analogous to the previous discussion of its antecedents, intelligence does not refer to a well defined set of subsequent responses, but to an elusive set of notions concerning problem solving. Postulating intelligence as the general something (force, energy, capacity, potential) which undergirds the abilities necessary to solve problems can, of course, neither be directly validated nor invalidated. However, by examining the consequences which have followed from this formulation, one may contrast its usefulness to that of other frames of reference not including intelligence in the above sense as the basis for understanding problem solving.

Intelligence as a General Factor

Analysis of intelligence as a psychological construct raises the problem of pointing to

the behavioral referents which are placed in the class, intelligence. The same analysis also requires determination of the grounds upon which certain aspects of behavior are classified. For, depending upon the basis of classification, the invented construct may fulfill its goal in expediting the understanding of behavior or as suggested by Peak (1953, p. 248) it may prove, "... artifactual and without systematic significance." Unfortunately it is not clear which are the behavioral referents for intelligence or which principles are involved in isolating the intelligence concept.

Intelligence as a prescientific concept (i.e., its meaning in everyday language) apparently refers to an unspecified but certainly very wide range of problem solving abilities. It is important to note that this usage does not refer directly to the behavior involved in solving problems but to the adequacy of the results of the behavior. In this sense the immediately distinguishable referents for intelligence are not behaviors but problems. The importance of all this lies in the recognition that selecting a sample of problems from an unknown universe of problems on some historical, cultural and/or common sense basis to measure intelligence (a procedure traditionally followed in intelligence test construction) may involve a number of behaviors which are not functionally related. However, the construct validity of intelligence as referring to all intellectual problems (i.e., a general factor) depends upon the functional relatedness of all or at least most of the behaviors involved in problem solving. That is, in terms of one set of operations, the describable events within the class, intelligence, must correlate as a function of a common source of variance.

Although factor analysis has repeatedly demonstrated the unrelatedness of a striking number of behaviors traditionally grouped under intelligence, there still persists the hypothesis of a general ability which now, as the result of training, interest, motivation, etc., has become differentiated into the various abilities demonstrated by correlational methods. The persistence of a general factor is based on the low but positive correlations often found between the factors which emerge from the test matrix (i.e., a second-order fac-

tor). However, even discounting similarities inherent in the content of the tests usually placed in the matrix, it cannot be assumed a priori that these correlations confirm the existence of general intelligence as usually understood. The communality which is found may, for example, be the result of the testing situation rather than any inherent property within the individual which transcends all situations. A series of studies conducted by Sarason and his coworkers (Gordon & Sarason, 1955; Mandler & Sarason, 1952; Mandler & Sarason, 1953; Sarason & Mandler, 1952) indicates that variables (test taking attitudes or test taking anxiety) operating as a function of the interpersonal competitive situation represented by the test may contribute significantly to test performance. It seems reasonable to hypothesize that these situational variables may to a large extent account for a general factor, and surely, this is not what is meant by intelligence. By systematically varying the conditions under which the same tests are given to the same individuals, it may be possible to demonstrate concomitant changes in the measures as a function of these situational variables.

Regardless of the accuracy of the hypothesis concerning the role of situational factors in accounting for G, the criticism of the existence of a general factor based on factor analytic methods stands. Guilford (1956, p. 290) in an extensive review of factor analytic studies of intellectual tests concludes that, "The methods of multiple factor analysis do not find a general psychological factor at the first order level and they find no second-order factor that can properly lay claim to the title of intelligence."

It is becoming increasingly clear that constructs which explicitly attempt to abstract functionally related behaviors in a situational class are potentially more useful than constructs which do not. For example, Liverant (1958) found increased predictability by dividing the need for recognition into the need for recognition in academic situations, and the need for recognition in social situations. It is reasonable to assume that greatly lowering the level of generality of intelligence by subcategorizing it in terms of specific classes of situations will significantly increase ac-

curacy of prediction. This likelihood has long been recognized within such categorization as social intelligence, academic intelligence, mechanical intelligence, etc. But somehow, we persistently return to a general intelligence, perhaps, because the meaning of the term proclaims the existence of one.

Cultural Definition

Whatever the most useful classification of intellectual behaviors proves to be, the very generality of intelligence makes it impossible actually to sample a unidimensional population of behaviors subsumed by intelligence since in one way or another most, if not all, of behavior sooner or later becomes involved in problem solving. Even scratching one's epidermis may be viewed as solving the problem of itching. However, such behavior tends not to be included in the typical intelligence test. In fact, the criterion of what is or what is not intelligent behavior involves a cultural value judgment. This makes for a great deal of difficulty in arriving at a consistent sample since by definition values are different in different cultures and tend to change with time in a given culture.

Rapaport (1953) suggests that in the case of terms having value connotations it is meaningless to ask what is X. Rather our problem becomes one of finding commonly accepted criteria for X. According to this point of view, rather than asking what is intelligence, we should be seeking established criteria for intelligent behavior which will vary from one class of situations to another. In this sense intelligence is divorced from its explanatory significance and serves a descriptive function, a state of affairs which may prove very desirable.

The Role of Nonintellective Variables

Reference to the situational context in which intellectual behaviors operate brings into focus the role of variables other than the intellectual behaviors per se in the problem solving process. While it is generally conceded that nonintellective factors are important in all actual measures of problem solving ability, within the intelligence model no clear conceptualization of what these variables are and how they function has yet been made. As

testers, we somehow assume that by holding nonintellective variables constant (vaguely referred to as establishing rapport or providing the proper conditions) during the testing period, an index of an individual's fundamental capacity to solve problems can be obtained apart from these nonintellective factors. But the manifestation of any problem solving ability in or out of the testing situation must be in accord with certain laws of behavior which include the operation of these so-called nonintellective variables.

An example using the terms of Rotter's (1954) Social Learning Theory may help clarify this latter point. In social learning theory the probability of the occurrence of any behavior depends upon the individual's expectation that using this behavior in the present situation will result in a certain kind of reward and the value of that reward for him. Thus, an individual may not manifest a certain implicit or explicit skill (e.g., memory, judgment, reasoning, etc.) necessary to solve a given problem for either of two reasons: one, the level of skill required is not within his repertoire; or, two, the right combination of expectancy and reward value necessary to elicit that behavior in that situation is not operating.

Neither intelligence tests nor the intelligence model provides an objective means of making a decision regarding which of the two reasons is operating in the case of failure to solve problems. Wechsler (1958) attempts to circumvent this problem by including the operation of nonintellective variables within his practical definition of intelligence. In practice, however, it is the inability to identify and formulate the operation of these variables which results in the lower predictive validity of intelligence tests, since it is these very variables which change as the situation changes. As a consequence of these situationally modifiable variables we have the fairly common case of the child who functions so well in school but behaves so "stupidly" with his peers or, what is even more puzzling within the confines of intelligence, does admirably on an intelligence test and abominably in school. It is only within a theory of problem solving which attempts an explicit formulation of all the variables necessary to predict behavior

that these apparent contradictions become understandable.

MENTAL DEFICIENCY

An adequate discussion of mental deficiency in light of the present suggestions concerning intelligence is beyond the scope of this presentation. It is mentioned at this point, however, because it is in relation to the intellectually handicapped that the discarding of intelligence as an explanatory construct may have its most far reaching effects. Fortunately, a number of prominent students (Hutt & Gibby, 1958; McCandless, 1952; Sarason, 1959) of this problem have anticipated and demonstrated the detrimental fallacy of regarding certain kinds of problem solving inadequacies as largely a matter of an inherited lack of capacity to perform any and all tasks at a certain level of difficulty. However, continued use of such terms as limited intelligence, inferior intellectual capacity, low IQ, feeble-minded, mental defective, etc. (all referring to lack of intelligence as an *explanation* for inadequate performance) hinders the understanding and the development of the point of view being expressed by these authors because of the surplus meanings inherent in the term intelligence. What is necessary from our point of view is recognition of the logical basis for discarding a contradictory and outmoded terminology which persists through tradition and not through theoretical and/or empirical justification.

AN APPROACH TO CHANGE

Inherent in the reasons for relegating intelligence to a nonexplanatory usage are suggestions for an alternative approach to understanding problem solving proficiency. Essentially what is being suggested is not that intelligence be replaced by a new term but that it be relegated to a descriptive function applicable to certain kinds of behavior whose explanation can be attempted by the continued development of existing constructs which can better integrate data of the kind previously mentioned and direct attention to new avenues of research. As argued by McCandless (1952) and Rotter (1958) among others, the constructs of modern learning

theories would appear to be a likely point of departure for this endeavor.

Rather than regarding intelligence tests as measures of a single dimension (general intelligence) it is more in keeping with the evidence to interpret them as assessing culturally expected standards of academic preparedness involving a number of complexly interrelated variables (e.g., implicit reaction patterns, motives, habits, expectancies, reinforcements, etc.) to account for the results. Attempting to incorporate the varieties of socially learned behaviors operating in the testing situation within the domain of developing theories of learning has the salutary effect of replacing a notion which is essentially untestable with constructs which have been selected with a view to functioning in potentially fruitful theories.

SUMMARY

An attempt has been made to question critically the validity of the term intelligence as an explanatory construct. It has been argued that the term intelligence carries with it essentially untestable and empirically unjustifiable connotations which preclude coherent and systematic organization of the awesome variety of activities ascribed to its domain.

An analysis of the antecedent and subsequent conditions defining intelligence reveals the following difficulties inherent in this concept:

1. The genetic model representing the antecedent conditions of intelligence is essentially untestable and not in accord with recent conceptual advances in both genetics and psychology.

2. The usefulness of postulating genetic antecedents for complex problem solving behaviors prior to a much more thorough understanding of how heredity and environment interact in a given case (a long and tortuous process at best) appears quite dubious both on theoretical and practical grounds. But more to the point are the more immediate possibilities of understanding and predicting intelligent behavior by continuing to pursue lines of investigation designed to reveal the situational conditions which facilitate or hinder the acquisition of various problem solving skills and the conditions which facili-

tate or hinder the expression of these skills in certain kinds of situations.

3. On the subsequent response side of the problem, the large majority of the evidence (primarily from factor analytic studies) indicates that a single dimension is not capable of subsuming the great varieties of behaviors now assumed to be related to intelligence.

4. The very generality of the term prohibits the isolation of a functionally related class of behaviors. Any attempt to establish a stable functional unity is further complicated by the definitional necessity of involving cultural value judgments in selecting a population of intelligent behaviors, judgments which vary from one culture to another and tend to change with time in a given culture.

5. Perhaps most crucial is the failure of intelligence theory to specify systematically the effect of situational variables on the behaviors it is purported to predict.

6. Finally, the arguments against the explanatory usage of intelligence suggest a fruitful approach to be the incorporation of the behavioral realm typically ascribed to intelligence within the confines of modern learning theory, particularly a social learning theory.

REFERENCES

- ANASTASI, ANNE. Some implications of cultural factors for test construction. *Proc., 1949 Conf. Test. Probl., Educ. Test Serv.*, 1950, 13-17.
- ANASTASI, ANNE. *Differential psychology*. New York: Macmillan, 1958. (a)
- ANASTASI, ANNE. Heredity, environment, and the question "how?" *Psychol. Rev.*, 1958, 65, 197-208. (b)
- ANDERSON, J. E. The prediction of terminal intelligence from infant and preschool tests. *39th Yearbook, Nat Soc. Educ.*, 1940, Part I, 385-403.
- AXLINE, VIRGINIA. Mental deficiency—symptom or disease. *J. consult. Psychol.*, 1949, 13, 313-327.
- BAYLEY, NANCY. On the growth of intelligence. *Amer. Psychologist*, 1955, 10, 805-818.
- BEACH, F. A., & JAYNES, J. Effects of early experience upon the behavior of animals. *Psychol. Bull.*, 1954, 51, 239-263.
- BEACH, F. A. Characteristics of masculine sex drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1956, 1-32.
- CARNAP, R. The methodological character of theoretical concepts. In H. Feigl & M. Scriven (Eds.), *Minnesota studies in the philosophy of science*. Vol. I. Minneapolis: Univer. Minnesota Press, 1956, 38-76.

- CLARKE, R. S., HERON, W., FETHERSTONHAUGH, M. L., FORGAYS, D. G., & HEBB, D. O. Individual differences in dogs: Preliminary report on the effects of early experience. *Canad. J. Psychol.*, 1951, 5, 150-156.
- COOMBS, A. W. Intelligence from a perceptual point of view. *J. abnorm. soc. Psychol.*, 1952, 47, 662-673.
- CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, 52, 281-302.
- EELS, K., DAVIS, A., HAVIGHURST, R. J., HERRICK, V. E., & TYLER, R. *Intelligence and cultural differences*. Chicago: Univer. Chicago Press, 1951.
- FROMM, E., & HARTMAN, L. *Intelligence: A dynamic approach*. New York: Doubleday, 1955.
- FULLER, J. L. *Nature and nurture: A modern synthesis*. New York: Doubleday, 1954.
- GORDON, E. M., & SARASON, S. B. The relationship between "test anxiety" and "other anxieties." *J. Pers.*, 1955, 23, 317-323.
- GUILFORD, J. P. The structure of intellect. *Psychol. Bull.*, 1956, 53, 267-293.
- HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, 56, 51-65.
- HUTT, M. L., & GIBBY, R. G. *The mentally retarded child*. Boston: Allyn & Bacon, 1958.
- HUXLEY, J. *Heredity east and west*. New York: Henry Shuman, 1949.
- JENNINGS, H. S. *The biological basis of human nature*. New York: Norton, 1930.
- JESSOR, R., & HAMMOND, K. R. Construct validity and the Taylor Anxiety Scale. *Psychol. Bull.*, 1947, 54, 161-170.
- JESSOR, R. The problem of reductionism in psychology. *Psychol. Rev.*, 1958, 65, 170-178.
- LIVERANT, S. The use of Rotter's Social Learning Theory in developing a personality inventory. *Psychol. Monogr.*, 1958, 72(2, Whole No. 445).
- MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 47, 166-173.
- MANDLER, G., & SARASON, S. B. The effect of prior experience and subjective failure on the evocation of test anxiety. *J. Pers.*, 1953, 21, 336-341.
- MCCANDLESS, B. Environment and intelligence. *Amer. J. ment. Defic.*, 1952, 56, 674-691.
- MCCLELLAND, D. C. *Personality*. New York: Sloane, 1951.
- MCGRAW, M. B. *Growth: A study of Johnny and Jimmy*. New York: Appleton-Century, 1935.
- PEAK, HELEN. Problems of objective observation. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953, 243-299.
- RAPAPORT, A. *Operational philosophy*. New York: Harper, 1953.
- ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
- ROTTER, J. B. A historical and theoretical analysis of some broad trends in clinical psychology. In S. Koch (Ed.), *Psychology: A study of science. Study II. Empirical substructure and relations with other sciences*. New York: McGraw-Hill, in press.
- SARASON, S. B. *Psychological problems in mental deficiency*. New York: Harper, 1959.
- SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, 47, 810-817.
- SHOFIELD, W. Critique of scatter and profile analysis of psychometric data. *J. clin. Psychol.*, 1952, 8, 17-22.
- SKINNER, B. F. Reinforcement today. *Amer. Psychologist*, 1958, 13, 94-99.
- SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Mental growth and personality development: A longitudinal study. *Monogr. Soc. Res. Child Developm.*, 1958, 23(No. 68).
- SPIKER, C. C., & MCCANDLESS, B. R. The concept of intelligence and the philosophy of science. *Psychol. Rev.*, 1954, 61, 255-266.
- STERN, C. *Principles of human genetics*. San Francisco: Freeman, 1952.
- STONE, J. A critique of studies of infant isolation. *Child Developm.*, 1954, 25, 9-20.
- WECHSLER, D. *The measurement and appraisal of adult intelligence*. Baltimore: Williams & Wilkins, 1958.

(Received May 18, 1959)

PERSONALITY DIFFERENCES AND LEVEL OF ASPIRATION

JUNE ELIZABETH CHANCE

University of North Carolina

A number of previous investigations have sought to study the relationship between personality differences and different patterns of level of aspiration. While many of these studies have found no stable relationships, a few (Cohen, 1954; Gruen, 1945; Klugman, 1948; Sears, 1940, 1941) find a recurrent relationship between "maladjustment" and particular patterns of goal setting. Specifically, two patterns appear to be more often characteristic of maladjusted subjects (Ss) than others, either very high (compensatory or wishful) or very low (protective) levels of aspiration. Cohen (1945), in discussing his results, asks what individual differences might explain why some Ss with strong feelings of self-rejection deal with the level of aspiration situation by making low, self-protective statements while others deal with the same situation by making high, unreal statements?

The present study concerns goal setting behavior as it relates to two independent dimensions of maladjustment in personality functioning. These two dimensions, designated A and R, were measured by two special MMPI scales devised by Welsh (1956). These scales, obtained by factor analytic methods, are uncorrelated with each other and may be used singly or in combination. While the scales are at present designated only as A and R, Welsh (1956) notes that high A scores are "related to disability of a dysthymic and dysphoric nature in which anxiety is prominent" while high R scores are "characterized by repression and denial." Preliminary work with psychiatric populations indicates that when the two scales are used together to classify Ss the resulting four groups are distinguishably different in psychiatric diagnosis (Welsh, 1956). Investigations with college students (Chance, 1957; Pepper, 1957; Saute, 1958) have like-

wise demonstrated differences in behavior between groups selected on the basis of position on both scales.

The present study was designed to investigate two hypotheses: (a) Ss scoring higher on A than on R (sensitizers)¹ will more characteristically employ a protective or low pattern of goal setting in a failure situation; (b) Ss higher on R than on A (repressors) will more often employ a compensatory or high pattern of goal setting when confronted with failure. These two hypotheses are consistent with a picture of the sensitizer as an individual who anticipates and fears failure and criticism and who deals with the situation by "admitting it first" and with a picture of the repressor as one who avoids discomfort by denying negative things about himself.

METHOD

Seven sections of elementary psychology courses were administered an adapted group form of the MMPI, containing the A and R scales, and a group level of aspiration measure during one of their regular class meetings. The group level of aspiration measure consisted of three sets of 15 anagrams and one set of 15 number series.

These four tasks were presented in booklet form with the following instructions:

On the following pages you will be asked to take a number of short experimental tests. These are new tests which we are just trying out. Preliminary work indicates that they may be good tests of intelligence or general academic aptitude. We are also interested in how well people can predict their own test performance before taking a test; therefore, before each test you will asked to make a prediction of the score you think you are most likely to obtain.

On each of the next three pages you will find 15 series of scrambled letters. If the letters are prop-

¹ The terms "sensitizers" and "repressors" will be used here for convenience. They refer only to the operational definitions given.

erly rearranged they will make a common English word. No proper names or trick words are included. Do any that you can; you need not work on them in order. You will have 2½ minutes to work on each set of anagrams.

When I say "Go" you may turn the page and begin. But before beginning, please write in the blank below the number of words you think you will be able to complete. On each of the series of anagrams the average college student is able to complete 8 of the 15 words. Nearly all students complete 4; a few students complete 12. Please do not turn the page until you are told to do so.

The instructions were read aloud by *E* while *Ss* followed them in their test booklets. *Ss* were asked to fill in identifying data on the front page as well as to make an estimate of their probable performance on the first set of anagrams. On a signal from *E*, all *Ss* turned to the second page and began working. They were told to stop at the end of 110 seconds. At the bottom of the second page they then recorded their estimates of their probable performances on the second set of anagrams. On a signal from *E* they turned to the third page and began work.

The procedure was repeated for the third and fourth pages; however, on the fourth page they were told:

On the next page you will find 15 number series. Four numbers are given and they are arranged according to some system. Figure out the system and then circle the one of the alternative numbers in the second column that fits the system. Three examples are given below.

The average college students gets 7 correct solutions out of 15 in 4½ minutes; nearly all complete 4, a few get 13. Please estimate in the blank below how many you think you will be able to complete. Do not turn the page until you are told to do so.

The time limit of 110 seconds for the sets of anagrams was so selected that the majority of *Ss*, if they set their aspiration close to the supposed group average of 8 correct solutions, could not succeed. (Although *Ss* were told that they were to be permitted a longer time to work than was actually allowed, no *S* questioned the imposed time limit.) For the number series they were actually allowed 4½ minutes to work. This longer time limit combined with the lower supposed "group average" was designed to let most *Ss* reach or exceed their aspiration estimate and hence to obtain some success at the end of the experiment.

Personality test and aspiration data were obtained from 147 *Ss*. In order to test the hypotheses regarding differences in aspiration behavior of sensitizers and repressors, *Ss* were selected from this larger group. Only male *Ss* were included in the experimental groups, since previous research has shown sex differences in level of aspiration behavior (Sheehan & Zelen, 1955; Sumner & Johnson, 1949).

A and *R* scores were translated into *T* scores by means of a table provided by Welsh (1956). Sensitizers were defined as those with *A* scores above 50, *R* scores below 50, and with at least a 10-point difference between the two scores. Repressors were similarly defined. For comparison purposes a control (*A* = *R*) group was also selected; controls obtained

TABLE 1
MEANS AND STANDARD DEVIATIONS OF PERFORMANCE SCORES, ASPIRATION ESTIMATES,
AND PERFORMANCE-ASPIRATION DIFFERENCES

Groups	Tasks							
	Anagrams 1		Anagrams 2		Anagrams 3		Number Series	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Sensitizers (<i>N</i> = 17)								
Performance	5.88	2.15	4.53	1.58	3.82	2.21	7.35	2.06
Aspiration	8.00	1.61	6.41	1.82	5.24	1.22	6.88	1.37
Difference	-2.12	2.57	-1.88	3.14	-1.42	2.44	—	—
Repressors (<i>N</i> = 25)								
Performance	5.44	1.65	4.08	1.92	4.52	1.72	7.08	2.50
Aspiration	8.48	1.58	7.32	2.05	6.52	2.10	7.12	1.68
Difference	-3.04	2.31	-3.24	2.72	-2.00	3.08	—	—
Controls (<i>N</i> = 26)								
Performance	6.15	2.19	4.65	1.99	4.69	2.72	7.42	2.70
Aspiration	8.46	1.43	7.15	1.59	6.73	1.52	7.27	2.41
Difference	-2.31	2.83	-2.50	2.07	-2.04	2.33	—	—

A and R scores within the range from 45 to 55 and had no more than 5 points difference between their two scores. In addition no S was included in any group who had actually succeeded in meeting or exceeding his aspiration estimate on any of the three anagram tasks.

RESULTS

Means and standard deviations of actual performances, aspiration estimates, and discrepancies between performance and aspiration on each task are given in Table 1. Mean differences between groups were evaluated by *t* tests. For each task (Anagrams 1, 2, and 3 and Number Series) the mean score of the sensitizers was compared with the mean score of the repressors; means of each of these groups were also compared with the mean of the control group. No differences between the groups in actual performance were significant. Likewise, the groups did not differ in size of discrepancies between performance and aspiration for the same task (which might be a measure indicative of intensity of failure). However, the mean aspiration estimate of the sensitizers for the third set of anagrams was significantly different from either the mean of the repressors or the mean of the control group. Values of *t* were, respectively, 2.25 ($p < .02 > .01$, *df* 40) and 3.43 ($p < .01 > .001$, *df* 41). None of the other means of aspiration estimates differed significantly among the groups.

Table 2 presents means and standard deviations of *d* scores, i.e., the discrepancy between aspiration estimate and performance on the previous trial. Mean differences were tested in the same manner as for the data in Table 1. None of the mean differences between groups

in the first *d* scores are significant, although the sensitizers had scores lower than the repressors as predicted ($t = 1.84$, $p < .10 > .05$, *df* 40). The mean difference between sensitizers and repressors in the second *d* score is highly significant ($t = 2.94$, $p < .01 > .001$, *df* 40). The same mean of the sensitizers approaches being significantly different from that of the control group ($t = 1.82$, $p < .10 > .05$, *df* 41). None of the comparisons of means of the third *d* score, which was based upon aspiration for the number series, were significant.

Examination of Tables 1 and 2 reveals that all groups begin the experiment with roughly equivalent levels of aspiration, perform in about the same manner, and fail to about the same extent. All groups tend to shift their aspiration estimates downward over the three trials of anagrams and to raise them again when the number series is introduced. However, the sensitizers as a group tend to shift their aspiration levels downward to a more marked degree than the other two groups, thus keeping the discrepancy between their estimate for any trial and their previous performance relatively small as compared with either the repressors or the control group. These data might be considered to confirm tentatively the hypothesis that sensitizers would be characterized by the use of a protective pattern of goal setting in a failure situation. The hypothesis that repressors would be characterized by a high, compensatory pattern fares less well. Although the repressors as a group do tend to keep the discrepancy between their last performance and next aspiration relatively larger than do the sensitizers, their *d* scores are not larger than those of an equal tendency control group.

DISCUSSION

The result, while suggesting the sensitizer is prone to adopt a protective pattern of goal setting in a failure situation, does not find any clear indication that the repressor adopts a compensatory pattern. Inspection of the data also failed to confirm an alternative hypothesis that repressors might be utilizing two or more patterns rather than just one. Both repressors and controls in making esti-

TABLE 2
MEANS AND STANDARD DEVIATIONS OF *d* SCORES
(ASPIRATION ESTIMATE MINUS PERFORMANCE
ON PREVIOUS TRIAL)

Groups	Scores					
	$d_1 = A_2 - P_1$		$d_2 = A_3 - P_2$		$d_3 = A_4 - P_3$	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Sensitizers	.53	2.53	.65	1.68	3.06	2.97
Repressors	1.96	2.39	2.44	2.06	2.60	1.77
Controls	1.00	2.47	2.08	1.64	2.58	2.98

mates either retained the estimate given for the previous trial or else lowered their old estimates slightly, i.e., there were no marked differences between repressors and controls in their responses to this experimental situation.

Although the original intention of the study was to use the A and R scales as two dimensions of maladjustment, the choice of operational definitions for sensitizers and repressors would argue that few, if any, Ss included in either group could legitimately be characterized as markedly deviant. Clearer differences between sensitizers and repressors and between both these groups and the control group might have been obtained if more extremely deviant Ss had been available.

In a carefully executed study, Miller (1951) investigated level of aspiration in psychiatric patients. Among his several criteria for inclusion of patients in one of the four diagnostic groups studied were patterns of scores from the clinical scales of the MMPI. Welsh's (1956) demonstrated relationships between the A-R classification and profile configurations on the MMPI suggests that the findings of the present study utilizing a "nondeviant" population may be profitably compared with Miller's findings from a "deviant" population. Miller found that neurasthenics produced on the average the highest positive *d* score, that both paranoid schizophrenics and a group of nonpsychiatric hospital patients produced *d* scores just slightly above zero, and that character disorders and conversion hysterics produced slightly negative *d* scores. Welsh found that elevation on *D*, *Pt*, and *Sc* were prominent in the profiles of high A/low R Ss and that psychiatric diagnoses of severe psychoneurosis, anxiety state, and paranoid schizophrenia occurred most often in this group. High R/low A Ss were more likely to show elevations on *Hs* and *Hy* and a marked lack of elevations on the psychotic scales; they were more likely in a psychiatric population to obtain diagnoses of mild psychoneurosis, hysteria, and conversion reaction. Examination of Miller's definition of neurasthenia (elevation on *Hs*, *D*, and to a lesser extent on *Hy*; lack of elevation on *P*, *Sc*, and *Ma*) indicates a fair degree of correspondence to Welsh's diagnostic expectations for the high R/low A group, and, as in the

present study, this is the group found to have the highest positive *d* scores. Similarly, Miller's diagnoses of character disorder and paranoid schizophrenia correspond rather clearly to the high A/low R scoring group described by Welsh. And again the observations of both studies are congruent; both groups show a low average *d* score, implying a self-protective orientation toward goal setting.

There are, however, some discrepancies in the findings of the two studies. Miller's normals set goals more like his paranoid schizophrenics and character disorders, while in the present results normals set goals more like repressors than sensitizers. The second departure in the findings is the negative *d* score of Miller's conversion hysterics. Following Welsh, we would expect these Ss to behave like the repressor group, obtaining positive rather than negative *d* scores. Miller's criteria for the diagnosis of conversion hysteria seem, however, to have been somewhat more stringent than those employed in the construction and validation of the *Hy* scale, which may account in part for the discrepancy of results. The over-all degree of correspondence between the present findings and Miller's lends considerable support to the general idea that characteristic patterns of goal settings and defending against failure are a significant aspect of different personality organizations all along the continuum of adjustment.

The present study also differs in several respects from many previous studies. One important difference is that Ss were subjected to consistent failure throughout the first three trials of the experiment; consequently, the obtained results could not be generalized to anything other than a failure situation. The influence of personality trends on goal setting might be expected to operate differently in situations where Ss succeed or where success and failure are interspersed. This experiment also utilized a relatively short series of trials; the character of the results might change if Ss were subjected to more prolonged failure.

SUMMARY

Groups of college students classified as "sensitizers" or "repressors" by means of the Welsh A and R scales (MMPI) were com-

pared in their performance on a group level of aspiration measure. The situation was arranged so that Ss experienced failure on the HOA tasks. Sensitizers employed a goal setting pattern closer to their real previous performance than did repressors or an equal tendency control group ($A = R$). Repressors did not show an hypothesized tendency to set high or compensatory goals. Limitations of the findings are discussed.

REFERENCES

- CHANCE, JUNE E. Some correlates of affective tone of early memories. *J. consult. Psychol.*, 1957, 21, 203-205.
- COHEN, L. D. Level of aspiration behavior and feelings of adequacy and self acceptance. *J. abnorm. soc. Psychol.*, 1954, 49, 84-86.
- GRUEN, EMILY W. Level of aspiration in relation to personality factors in adolescents. *Child Develpm.*, 1945, 16, 181-188.
- KLUGMAN, S. F. Emotional stability and level of aspiration. *J. gen. Psychol.*, 1948, 38, 101-119.
- PEPPER, L. J. The relationship between Welsh Figure Preference Test responses and indices of anxiety and repression. Unpublished master's thesis, Univer. of North Carolina, 1957.
- MILLER, D. R. Responses of psychiatric patients to threat of failure. *J. abnorm. soc. Psychol.*, 1951, 46, 378-387.
- SAUTE, LOIS C. Judge personality and favorability biases in interpersonal predictions. Unpublished doctoral dissertation, Univer. of North Carolina, 1958.
- SEARS, PAULINE S. Levels of aspiration in academically successful and unsuccessful children. *J. abnorm. soc. Psychol.*, 1940, 35, 498-536.
- SEARS, PAULINE S. Level of aspiration in relation to some variables of personality: Clinical studies. *J. soc. Psychol.*, 1941, 14, 311-336.
- SHEEHAN, J. G., & ZELEN, S. L. Level of aspiration in stutterers and nonstutterers. *J. abnorm. soc. Psychol.*, 1955, 51, 83-86.
- SUMNER, F. C., & JOHNSON, E. E. Sex differences in levels of aspiration and self estimates of performance in a classroom situation. *J. Psychol.*, 1949, 27, 483-490.
- WELSH, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.

(Received February 13, 1959)

MANIFESTATIONS OF HOSTILITY IN NEURODERMATITIS¹

M. MICHAEL KLABER²

Teachers College, Columbia University

Neurodermatitis is a relatively common chronic skin disorder, characterized by intensely pruritic excoriations, and is usually attributed to psychogenic factors. The nature of these psychological disturbances is described as centering around the suppression of intense hostile impulses (Dunbar, 1954; Obermayer, 1955). Neurodermatitis patients thus present an opportunity to investigate a specific personality pattern in a relatively objectively classified population. The purpose of this research was to test by an experiment the conclusions obtained through clinical observations, and to contribute to a better understanding of the disorder.

Early studies using rating scales and personality inventories (Allerhand, Gough, & Grais, 1950; Greenhill & Finesinger, 1942; Lynch, Hinckley, & Cowan, 1945) unanimously reported a higher incidence of hostility in neurodermatitis patients than in various control groups. None of these investigators, however, reported a cross-validation of their findings.

Several projective studies are reported in the literature (Cleveland & Fisher, 1956; Fiske, 1953; Fiske & Obermayer, 1954; Geist, 1957; Kepecs, Rabin, & Robin, 1951; Levy, 1952). All investigators concurred with the finding that neurodermatitis patients manifest more hostility than do normal persons on projective tasks. Little is known,

however, of the relationship between covert hostility as manifested in projective protocols and overt expression of aggression. In an effort to correlate various levels of assessment of behavior, Leary (1957, pp. 395-396) constructed a multilevel personality battery. Reporting on skin patients he described his findings: "The multilevel pattern is that of a sweet, responsible facade with underlying rage and bitterness . . . they exhibit an intense conflict between this bland exterior and sadistic rage." Unfortunately, no control group was used for comparison with the skin patients.

An outstanding therapeutic experiment conducted by Seitz (1953) emphasized the suppression of hostility in dermatitis patients and based a therapeutic procedure on this characteristic. Reporting on a series of 25 patients, he arrived at the conclusion that the alleviation of skin eruptions could be achieved through a short series of therapeutic interviews focusing only on hostility and aggressive impulses. In a period of 10 sessions he explored possible sources of annoyance to the patient and encouraged him to talk about them. Only 13 of the 25 patients remained in therapy; 12 of the 13 were completely relieved of symptoms.

Summarizing the present state of knowledge, the neurodermatitis patient may be described as a generally hostile person who expends great effort on successful control of his impulses. Outwardly he appears calm and collected, yet covertly he harbors deep resentment. In psychotherapy such a patient will readily respond to favorable interpersonal conditions by verbalizing a great deal of hostility. This research was designed to test these formulations.

¹ This study is based on a doctoral dissertation submitted in partial fulfillment of the requirements for the PhD degree at Teachers College, Columbia University. Thanks are extended to Joel R. Davitz, Paul E. Eiserer, and especially to Laurance F. Shaffer, who was of invaluable assistance during all stages of this study. The author is indebted to A. C. Carr and G. S. Baroff for serving as judges.

² Now at the Institute for the Crippled and Disabled.

METHOD

The instruments were chosen to reflect the central issues of the research: They were (a) a Manifest Scale of Hostility to test the hypothesis that neurodermatitis patients do not see themselves as more aggressive than normal persons; (b) seven Thematic Apperception Test (TAT) cards selected to elicit hostility and thus to tap the covert level of this specific behavior; and (c) a Sentence Formation Task (SFT) on which sentences were constructed and which could be reinforced. It was postulated that the SFT best reflects the therapeutic situation and thus represented a laboratory analogue to Seitz's method.

The Manifest Scale of Hostility

A manifest scale on which the subject (S) evaluated himself with respect to hostile behavior was chosen for comparing overt hostility. Inasmuch as the *Ne* scale (Allerhand et al., 1950) had been especially designed for use with neurodermatitis patients and contained a number of items dealing with the central issue of this research, its use seemed appropriate.

Twenty-four items from the *Ne* scale were included in the instrument, supplemented by four additional items in order to obtain greater reliability. The score was the sum of the items judged by two psychologists to pertain to hostility. In such a manner 15 items were included in the final score.³ In spite of the added items in an effort to obtain a reliable measure, the index of split-half reliability was low ($r = .39$).

The Thematic Apperception Test

For the purpose of obtaining a covert hostility score, a series of seven TAT cards was administered in the standard manner. Cards 3 BM, 6 BM, 8 BM, 12 M, 13 MF, 18 GF, and 18 BM were used. The stories obtained were independently analyzed by two clinical psychologists according to the scoring instructions developed by Fisher and Hinds (1951). It was emphasized that scoring should not involve interpretation, but mere classification of content. Seven categories of hostility were classified from story content: (a) openly hostile stories, (b) verbally expressed hostility, (c) rationalized hostility and suicide, (d) illness and accidents, (e) ambivalently expressed hostility, (f) neutral content, and (g) posited hostility or favorable content. Interjudge reliability for the scoring categories ranged from .79 to .85 on the different scales. In those instances where judges disagreed, a third clinician was consulted to resolve the issue.

The scores were weighted in the following manner:

³ After the inception of the research Buss and Durkee published a Hostility-Guilt inventory which may have been particularly useful in this study. The publication came too late, however, to be incorporated in the research design (Buss & Durkee, 1957).

a score of 3 was assigned to Category *a* (openly hostile stories), a score of 2 to Categories *b* and *c* (verbal expressions of hostility and rationalized hostility), a score of 1 to Categories *d* and *e* (illness, accidents, and ambivalently expressed hostility), a score of -1 to neutral content (Category *f*), and of -2 to positive content (Category *g*). The final score was the algebraic sum of the scores of the seven stories for one individual.

The Sentence Formation Task

While some previous studies used overt and covert measures, few, if any, attempted to measure the process of psychotherapy itself. Psychotherapy has been viewed as a process of selective reinforcement by the therapist (e.g., Shaffer, 1947). Inasmuch as the elicitation of hostile responses is central to the psychologic treatment of neurodermatitis patients, it was believed that verbal reinforcement of hostile responses is analogous to the therapeutic situation. A recent review by Krasner (1958) summarized the available research in this area.

In order to elicit reinforcing responses, a sentence completion task was constructed. On each of 74 3" x 5" cards four verbs were typed, one of which was of hostile stimulus value. The S was presented with one card at a time and instructed to construct one sentence using any one of the four verbs presented to him.

The hostility value of a verb was determined by two clinical psychologists. Each was given a list of 82 verbs selected for possible hostility value, and was asked to indicate which of the verbs would be termed "hostile" when used in a sentence. Both judges agreed on 74 verbs, and these were used in the construction of the instrument.

Three scores were obtained for each individual consisting of the number of verbs of hostile stimulus used in each of the three conditioning periods: operant, reinforced, and extinction. A split-half reliability coefficient was computed for each of the three experimental conditions. The correlations for hostile responses were .60, .62, and .77, respectively.

Procedure

The manifest scale was administered in a mimeographed form on which the patient had to indicate, by circling either T or F, whether a statement applied to him or not.

The seven TAT cards were administered in the standard manner.

The Sentence Formation Task was introduced with the following statement: "I want to see how good you are at making sentences. I am going to show you some cards, each of which has four verbs. You can choose to make a sentence with any of the verbs." The first 30 cards were presented without any comment in order to establish an operant level of response. On the next 30 cards, whenever the S chose the hostile verb to make a sentence, he was reinforced by the experimenter who said, "yes," "good," "fine," "uh-huh," or "hmm-mm," as the oc-

casion warranted. The presentation of the last 14 cards was regarded as an extinction period and no reinforcement was administered. The cards were presented in a random order.

The testing session required from 45 to 60 minutes.

Subjects

The Ss of the investigation were 40 patients referred through the Department of Dermatology of the Presbyterian Hospital and the College of Physicians and Surgeons of Columbia University as well as from one private dermatologist.⁴ Three experimental Ss were obtained through the latter source, while the rest of the neurodermatitis patients, as well as all the controls, through the former.

In order to control for the psychological effects of skin disorder per se, all control Ss were derived from the same general population as the experimental group, except for their diagnosis. Only patients afflicted with clearly nonpsychogenic disorders (e.g., contact dermatitis, lupus, etc.) were accepted. The experimental Ss were diagnosed independently by a board-certified dermatologist or by the staff of the clinic.

All Ss were white, over 18 years of age, and were born or had received their schooling in this country. The experimental group consisted of 5 males and 15 females, and the control Ss were 8 males and 12 females. No significant differences were found between the experimental and control groups with respect to age (the means were 38.75 and 38.10, respectively) and years of school attended (means 12.35 and 12.85, respectively).

RESULTS AND CONCLUSIONS

Overt Hostility

No significant differences between the neurodermatitis patients and the control group were found ($t = 1.02$ with 38 *df*). One must bear in mind, however, that the reliability coefficient of the scale used was only .39, so that the acceptance of the null hypothesis applied only to the measure used. Further research with a more reliable scale would throw additional light on the matter at hand.

Covert Hostility

It had been hypothesized that neurodermatitis patients show greater hostility than the control group on instruments tapping covert personality factors, such as measured by the

⁴ The author is indebted to Carl T. Nelson, Executive Officer of the Department of Dermatology, for his permission to use the Ss and for his active cooperation in the study, and to D. P. Torre and J. McCarthy for their assistance and cooperation.

TABLE 1
COMPARISONS OF MEANS AND STANDARD DEVIATIONS ON TAT SCORES

Statistic	ND Patients	Control	Statistics of Comparison	<i>p</i>
Mean	6.7	3.3	$t = 2.83$.01
SD	5.1	2.8	$F = 3.22$.02

TAT. This hypothesis is borne out by the results of the research. Table 1 shows both significantly higher mean hostility scores and greater variance for the experimental Ss in comparison with the dermatology control group. These results show that neurodermatitis patients emit both more hostile responses and are as a group more variable on a projective test designed to tap covert personality variables.

Response to Reinforcement

From the reports of psychotherapists, it became evident that neurodermatitis patients respond favorably to verbal reinforcement of hostile responses. It was therefore hypothesized that such patients manifest different response frequencies of verbal hostile behavior under experimental reinforcement conditions than normal persons.

Three measures were computed for purposes of inter- and intragroup comparisons,

TABLE 2
COMPARISONS OF MEANS AND STANDARD DEVIATIONS OF HOSTILITY SCORES ON SFT
(In percent possible score)

Experimental Condition	ND Patients	Controls	Statistics of Comparison	<i>p</i>
Operant				
Mean	10.8	14.3	$t = 2.04$.05
SD	5.0	5.6	$F = 1.26$	NS
Reinforced				
Mean	18.2	18.5	$t = 0.83$	NS
SD	9.4	8.6	$F = 1.19$	NS
Extinction				
Mean	18.5	18.5	$t = 0.00$	NS
SD	9.8	9.2	$F = 1.13$	NS

corresponding to operant, conditioning, and extinction periods. As can be seen from Table 2, neurodermatitis patients show a significantly lower hostile response frequency on the SFT under nonreinforced conditions than control Ss. Under favorable (i.e., reinforced) conditions, however, they emit no fewer hostile responses than the control group.

As it was originally stated, the hypothesis with respect to increased verbal response rates of neurodermatitis patients under favorable conditions was not substantiated. These patients differ from the control group not in response to reinforcement, but in their base level of verbal hostility. Reinforcement enables these persons to achieve a normal rate of emission of verbal hostile responses. The results are consistent with the clinical observations.

Hostility and Neurodermatitis

Certain conclusions with respect to the manifestations of hostility in neurodermatitis patients may be drawn from the test battery used in this research. Assuming that the Manifest Scale of Hostility measures overt, and the TAT covert, expressions of aggressive impulses, the neurodermatitis patients emerge as persons with a covert hostility drive which they do not manifest overtly. This conclusion is substantiated by their low operant level of emittance of hostile verbs. The spontaneous, unreinforced response may constitute a sample of the verbal hostility rate of the person in everyday life. A low operant level thus reflects infrequent use of hostile words and consequently placid verbal behavior.

DISCUSSION

The results of this study demonstrate the feasibility of constructing a test battery specifically designed to measure overt and covert levels of a specific personality dimension, as well as the verbal emission of certain response categories. Such batteries may, in the future, be used to investigate the nature of the psychotherapeutic process and may contribute to a better understanding of the verbal interaction between therapist and client.

The research study concerned itself with a specific medical-diagnostic group of patients

and confirmed impressions formulated by clinical workers with respect to one character dimension (hostility), but was not designed to test all hypotheses concerning the etiology of the disorder. It is entirely possible, for example, that other personality characteristics, such as anxiety, are concomitants of the neurodermatitis syndrome. Furthermore, it is possible and quite likely that disturbances in the psychologic sphere are necessary contributing factors to the etiology of the disorder, but that an organic predisposition constitutes a prerequisite for the formation of the syndrome.

The findings of the study are limited by the deficiencies of the instruments used. The Manifest Scale of Hostility in particular was found to be so low in reliability that the acceptance of the null hypothesis with respect to the two groups may well have resulted from the inadequacy of the scale rather than from a real absence of difference in response pattern. Inasmuch as available research presents conflicting evidence with respect to overt and covert relationships, the data of this study are insufficient to demonstrate conclusively that the hostility manifested in neurodermatitis patients is demonstrable on a purely covert level. Taken at face value, the results obtained on the TAT and the SFT are consistent with theories postulating an inverse relationship with respect to overt and covert hostility patterns. The lower unreinforced use of hostile verbs on the latter measure can be regarded as a behavioral measure of verbal hostility.

The experimental procedure in the research was designed in such a manner as to permit future applications to research with the process of psychotherapy. The SFT appears suited to simulate an experimental psychotherapeutic relationship.

SUMMARY

Substantial clinical evidence suggests that patients afflicted with neurodermatitis suffer from a personality disturbance in which the suppression of intense hostility is a central feature. This research was designed to obtain empirical evidence concerning the manifestations of hostility on overt and covert levels in such patients.

It was hypothesized that neurodermatitis patients (a) evidence no more hostility than control Ss suffering from nonpsychosomatic skin diseases on an instrument tapping overt behavior, (b) that neurodermatitis patients manifest more hostility on a covert measure of hostility than the control Ss, and, finally, (c) that neurodermatitis patients differ in verbal hostile responses on a conditioning task.

The results were consistent with all three hypotheses. No difference between the two groups was found on the Manifest Scale of Hostility, but the low reliability of the scale does not permit more than a tentative acceptance of the null hypothesis. Differences on the TAT, as rated by two independent clinicians, were significant at the .01 level with respect to both higher mean and larger variance in favor of the neurodermatitis group. On the Sentence Formation Task the control patients emitted significantly more hostile responses on the operant base level than the experimental Ss. The experimental Ss increased their hostile responses under reinforcement conditions, reaching a level equal to that of the control Ss. Some of the implications of the methodology with respect to future research were discussed.

REFERENCES

- ALLERHAND, M. E., GOUGH, H. G., & GRAIS, M. C. Personality factors in neurodermatitis: A preliminary report. *Psychosom. Med.*, 1950, 12, 387-390.
- BUSS, A. H., & DURKEE, ANN. An inventory for assessing different kinds of hostility. *J. consult. Psychol.*, 1957, 21, 343-349.

- CLEVELAND, S. E., & FISHER, S. Psychological factors in the neurodermatoses. *Psychosom. Med.*, 1956, 18, 209-220.
- DUNBAR, FLANDERS. *Emotions and bodily changes*. (4th ed.) New York: Columbia Univer. Press, 1954.
- FISHER, S., & HINDS, EDITH. The organization of hostility controls in various personality structures. *Genet. psychol. Monogr.*, 1951, 44, 3-68.
- FISKE, C. E. Personality and emotional factors in chronic disseminated neurodermatitis. Unpublished doctoral dissertation, Univer. of California, Los Angeles, 1953.
- FISKE, C. E., & OBERMAYER, M. E. Personality and emotional factors in chronic neurodermatitis. *Arch. Derm. Syph., N. Y.*, 1954, 61, 261-268.
- GEIST, H. Emotional aspects of dermatitis. *J. clin. exp. Psychopath.*, 1957, 18, 87-93.
- GREENHILL, M. H., & FINESINGER, J. E. Neurotic symptoms and emotional factors in atopic dermatitis. *Arch. Derm. Syph., N. Y.*, 1942, 46, 187-200.
- KEPECS, J. G., RABIN, A., & ROBIN, M. Atopic dermatitis. *Psychosom. Med.*, 1951, 13, 1-9.
- KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, 55, 148-170.
- LEARY, T. *Interpersonal diagnosis of personality: A functional theory of personality evaluation*. New York: Ronald, 1957.
- LEVY, RUTH. Rorschach pattern in neurodermatitis. *Psychosom. Med.*, 1952, 14, 41-49.
- LYNCH, F. W., HINCKLEY, R. G., & COWAN, D. W. Psychosomatic studies in dermatology: B. Psychobiologic studies of patients with atopic eczema (disseminated neurodermatitis). *Arch. Derm. Syph., N. Y.*, 1945, 51, 251-257.
- OBERMAYER, M. E. *Psychocutaneous medicine*. Springfield, Ill.: Charles C Thomas, 1955.
- SHAFFER, L. F. The problem of psychotherapy. *Amer. Psychologist*, 1947, 2, 459-467.
- SEITZ, P. F. Dynamically-oriented brief psychotherapy: Psychocutaneous excoriation syndromes. *Psychosom. Med.*, 1953, 15, 200-242.

(Received February 16, 1959)

A COMPARISON OF CHANGES IN PSYCHONEUROTIC PATIENTS DURING MATCHED PERIODS OF THERAPY AND NO THERAPY¹

ROSALIND DYMOND CARTWRIGHT AND JOHN L. VOGEL²

University of Chicago

Despite a very large increase in the number of published research reports on the effects of psychotherapy, the more basic question as to its effectiveness seems still to remain unanswered. This question has been framed most often as one of the effectiveness of therapy versus spontaneous remission. It has been dealt with in these terms in a review by Eysenck (1952) and in recent research studies by Barron and Leary (1955), and Dymond (1955). Barron and Leary concluded from their careful study of MMPI changes that therapy patients did not improve significantly more than did waiting list controls. Cartwright (1956), in a note on their paper, pointed out that though therapy patients did not *improve* more, there was significantly more *change* with regular, formally defined psychotherapy than without it. This was indicated by the difference in the variances of the changes for the treated and untreated groups. Dymond (1955) reported a comparison of improvements made by a small sample of waiting list "improvers" and therapy "improvers." The analysis strongly suggested the improvement of the two groups to be different in kind.

Although neither study gives any definitive answer to the basic question, each supplies some evidence of therapy having some measurable impact not found in its absence. The study to be reported here is another of this kind: it appears to help clarify the argument that therapy does have some special effectiveness with psychoneurotic patients by specifying

ing the conditions of the effectiveness more fully.

HYPOTHESES

The major hypothesis can be stated most generally as:

1. There will be more change in measured adjustment for a self-referred group of psychoneurotic patients during a period of time in which they are meeting regularly with a therapist than during a period of time equal to it during which no regular formally constituted therapy is taking place.

Several subhypotheses will also be investigated. If it is supposed that therapy has more effect than no therapy, then:

2. Longer therapy periods will be associated with more improvement than shorter therapy periods.

3. Longer waiting time periods will not be associated with more improvement than shorter waiting periods.

4. Patients seen by experienced therapists will improve more in a therapy period than those seen by inexperienced therapists.

METHOD

Design and Subjects

Thirty subjects (Ss) who applied to the University of Chicago Counseling Center for therapy were asked to participate in a research study which would involve testing at four points in time: (a) immediately upon their agreement to participate (Pre-wait test), (b) immediately before beginning counseling (Pre-therapy test), (c) sometime during the course of their therapy at a point exactly equal in time to the length of time between Test 1 and Test 2 (In-therapy test), and (d) when therapy had been completed (Post-therapy test).

Of the 30 Ss who were initially tested, 8 remained

¹ This investigation was supported by a research grant from the Ford Foundation to the University of Chicago Counseling Center.

² Now at University of Washington.

TABLE 4

EFFECT OF PRE-WAIT STATUS ON DIRECTION OF CHANGE AND LENGTH OF WAIT (TAT)

(Direction of Change on TAT while Waiting)

	+	0-	Total
High 10+			
Pre-Wait status			
Long Waits	3	3	6
8-24 Weeks			
Short Waits	1	4	5
3-7 Weeks			
Low 9-			
Pre-Wait status			
Long Waits	6	0	6
Short Waits	2	3	5
	12	10	22

the three variables—length of time (waiting or In-therapy), original status, and direction of change—a two-by-two-by-two analysis was used.⁶ The first of these was designed to answer the question "Does the original Pre-wait status on the TAT affect the association between the length of waiting time and direction of change on this measure?" Separate two-by-two tables were set up to look at the direction of change positive as against no change or negative change for long wait and short wait Ss, one for those whose original status on the TAT was high and the second for those whose Pre-wait status was low.

These two tables were then combined in terms of number of cases on the diagonal and number off the diagonal. The difference between the proportions on the diagonal for high and low original status was tested for significance using the Fisher and Yates exact method. In this case, the difference between the proportions 7/11 and 9/11 is not significant. The answer to the original question, then, is that long wait Ss improve and short wait Ss do not improve or deteriorate on the TAT irrespective of their original status.

A second analysis of this type was done to determine if deterioration of the high Pre-therapy TAT scorers and improvement of the

TABLE 5

CONDENSATION OF TABLE 4

	On-Diagonal	Off-Diagonal	Total
High	7	4	11
Pre-Wait status			
Low	9	2	11
Pre-Wait status			
	16	6	22

low Pre-therapy scorers during a therapy period was affected by the length of the time in therapy. The proportion on the diagonal for the long therapy group in this case was 7/10 and not significantly different from the proportion on the diagonal for the short therapy group 9/12. Thus the relationship found in Table 3 holds: those with high TAT scores before therapy begins tend to deteriorate or show no change, while those with initially low scores tend to improve these scores regardless of the length of time involved in the In-therapy period. It can now be stated in relation to Hypothesis 2 that longer therapy is not related to more improvement on either test. Hypothesis 3 that longer waiting would not be related to more improvement is confirmed for the Q adjustment score, but not for the TAT, where longer waiting was found related to more improvement regardless of original status.

Testing Hypothesis 4

Hypothesis 4 states that Ss seen by experienced therapists will improve more on the

TABLE 6

RELATION OF CHANGE IN THERAPY TO EXPERIENCE OF THERAPIST

(Change while in therapy)

	Q Adjustment		TAT Score	
	+	0-	+	0-
Exper.	6	6	8	4
Inexper.	7	3	1	9
	NS $\chi^2 = 7.245^*$			

⁶ This technique was suggested by William Kruskal, to whom the authors wish to express their thanks.

* Significant at .01 level.

two test measures than those seen by inexperienced therapists. To determine if there were any bias in the selection of the cases seen by those two groups, the direction of change while waiting was examined. The experience of the therapist-to-be-seen was found to be unrelated to changes before therapy begins. Table 6 shows the relation of the direction of change while *in therapy* to the experience of the therapist. Before the very high degree of significance in Table 6 is taken as clear evidence of the positive effect of experienced therapists on the TAT changes of their patients, several other factors must be explored for their possible interaction.

1. Since length of In-therapy time showed a nonsignificant trend (.10) in the direction of longer cases changing negatively on the TAT, the first question of this kind is: are the experienced therapists seeing more cases that waited a shorter time to begin therapy and so had a shorter In-therapy period and a bias toward positive change? This was found to be a nonsignificant relationship in a chi square test.

2. Since Pre-therapy status on the TAT was found to be negatively related to In-therapy direction of change (see Table 3), are

the experienced therapists perhaps seeing more cases with initially low scores? This, too, was found not to be the case using a chi square.

An on-diagonal, off-diagonal analysis confirmed these findings. The relation of the experience of the therapist to the direction of the TAT change was not significantly affected by the Pre-therapy status nor by the length of time in therapy. It seems clear that the experience of the therapist is the most powerful influence of those investigated here in producing positive change in the patient's TAT mental health rating during therapy.

Up to this point then, it can be said that there were two conditions under which the TAT mental health rating improved significantly: (a) without therapy, where the waiting period before beginning therapy was 8 to 24 weeks (Long-wait), and (b) with therapy, where the patient was seen by an experienced therapist. As far as the TAT is concerned, it seems a reasonable conclusion that over the time periods involved here, no therapist at all was better than an inexperienced one. Whether the gains made while waiting would have been sustained if no therapy had been entered into remains an open question, but a comparison of the mean changes under these

TABLE 7
MEAN SCORES FOR EXPERIENCED AND INEXPERIENCED GROUPS

	Pre-Wait	Pre-Therapy	In-Therapy	Post-Therapy	<i>t</i> Pre-In	<i>t</i> Pre-Post
Q Scores						
Exper.			41.0		—	
<i>N</i> = 12	39.2	39.7		46.0		2.536*
<i>N</i> = 11		39.0				
Inexper.			37.3		2.163*	
<i>N</i> = 10	33.1	31.0		40.25		1.563
<i>N</i> = 8		32.81				
TAT Scores						
Exper.			12.58		2.98**	
<i>N</i> = 12	9.92	10.41		12.72		2.190*
<i>N</i> = 11		10.27				
Inexper.			11.40		—	
<i>N</i> = 10	10.30	12.80		10.25		1.807
<i>N</i> = 8		12.87				

* Significant at .05.

** Significant at .02.

conditions confirms the above conclusion. The mean waiting period improvement for the Long-wait group was +2.9. The mean in-therapy time improvement for those seen by experienced therapists was +2.0. The t between these, .677, was not significant. The mean In-therapy change for those seen by inexperienced therapists was -1.4. The t between this and the improvement of the Long-wait group is 3.014, significant beyond the .01 level. Is the conclusion, then, that there is no difference between spontaneous remission and therapy change brought about by an experienced therapist and that inexperienced therapists have, if anything, a deleterious effect on their patients? As this conclusion would have strong implications, it is important that it be explored carefully. For this reason, the analysis was extended to the post-test data.

Table 7 reports the mean scores for both tests at the four test points for those seen by the two groups of therapists. A t test analysis for the differences between these means was used. The significant improvement between the Pre-therapy and the In-therapy test on the TAT of those in the experienced therapists' group has already been noted. A further result appeared to be a significant improvement on the Q adjustment score during this time of those in the *inexperienced* therapists' group. However, when an on-diagonal, off-diagonal analysis was used to examine the relationship of Q change in therapy and therapist experience for high and low Pre-therapy Ss, it was found that these proportions (8/11 on diagonal for low scorers) was significantly different from the proportion (4/11 off-diagonal for high scorers) at the .05 level. Thus, the improvement on Q adjustment for Ss in therapy with inexperienced therapists is largely a function of bias in Pre-therapy status. The N s for the comparisons between Pre- and Post-therapy scores were reduced to 19 by the loss of 3 cases; 1 from the experienced and 2 from the inexperienced group. The results here show that patients seen by experienced therapists improve significantly by the end of therapy on both measures, whereas those seen by inexperienced therapists failed to improve significantly on the Q score and, in fact, came close to a significant deterioration on the

TAT. As has been noted, those who waited a long time improved significantly on the TAT, but there was no commensurate shift of the Q adjustment score.

One difference between improvement with and without therapy has been found in this study. Those who waited a long time before beginning improved on the projective measure of their mental health but did not incorporate this into their self-picture as reflected by the Q score. Those who completed therapy with experienced therapists reflected their "deep" improvement on the conscious level in an improved self-description. Those completing therapy with inexperienced therapists not only did not improve their self-picture, but bordered on a deterioration in health.

SUMMARY AND CONCLUSIONS

Twenty-two self-referred psychoneurotic patients were tested at four points in time: (a) Pre-wait—on first being accepted as therapy cases and placed on the waiting list; (b) Pre-therapy—after a waiting time which varied in length; (c) In-therapy—after therapy had begun at a point equal to the waiting interval; and (d) Post-therapy. The instruments employed were the (a) Butler and Haigh Q sort, a self-descriptive test from which the Q adjustment score was computed, and (b) the TAT, a projective, from which a diagnostic rating was made. Four hypotheses were tested:

1. There will be more change during a period of therapy than an equal period of no therapy. This was confirmed for the Q adjustment score but not for the TAT.

2. Longer periods of therapy will be associated with more improvement than shorter periods. This hypothesis was not confirmed for either test.

3. Longer waiting periods will not be associated with more improvement than shorter periods. This was confirmed for the Q adjustment score but not for the TAT. Those who waited longer did in fact improve on the TAT, and those with shorter waits did not. The long wait improvers appear to be a spontaneous remission group with respect to the TAT.

4. The fourth hypothesis that patients in therapy with experienced therapists would improve more in the In-therapy period than

those in therapy with inexperienced therapists was confirmed for the TAT but not for the Q score. The fact that the positive effect of the experienced therapists was clearly independent of the Pre-therapy level of the TAT score and of the length of time in therapy made a strong case for the special effectiveness of these therapists. This in-therapy improvement could not be attributed to a continuation of a spontaneous remission trend.

The Post-therapy tests showed those in therapy with experienced therapists to have improved significantly on both tests, whereas those in therapy with inexperienced therapists not to have improved significantly on either and, in fact, they bordered on a significant decrease in health on the TAT.

It would appear that this study could conclude:

1. Therapy affects adjustment based on self-description, whereas equal waiting time does not.

2. A long waiting period for this type of patient improves the mental health picture as revealed by the TAT.

3. Equally long therapy does not necessarily improve the mental health picture.

4. An experienced therapist is effective in improving the mental health both as revealed by the TAT and as self-described.

5. An inexperienced therapist is ineffective in improving the self-picture and may actu-

ally be effective in decreasing the health as revealed by the TAT.

6. From the last two conclusions, it appears that therapists have special effects, depending on the level of their experience.

REFERENCES

- BARRON, F., & LEARY, T. Changes in psychoneurotic patients with and without psychotherapy. *J. consult. Psychol.*, 1955, 19, 239-245.
- BUTLER, J., & HAIGH, G. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 55-75.
- CARTWRIGHT, D. S. Note on "Changes in psychoneurotic patients with and without psychotherapy." *J. consult. Psychol.*, 1956, 20, 403-404.
- DYMOND, ROSALIND. Adjustment changes over therapy from Thematic Apperception Test ratings. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 109-120. (a)
- DYMOND, ROSALIND. Adjustment changes over therapy from self-sorts. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 76-84. (b)
- DYMOND, ROSALIND. Adjustment changes in the absence of psychotherapy. *J. consult. Psychol.*, 1955, 19, 103-107.
- EYSENCK, H. The effects of psychotherapy: An evaluation. *J. consult. Psychol.*, 1952, 16, 319-324.

(Received March 17, 1959)

TOWARD THE ELIMINATION OF THE CONCEPT OF NORMALITY¹

DAVID FREIDES

Lafayette Clinic, Detroit, Michigan

In recent years, the literature concerning the nature of normality in human personality has shown increasing support for an "idealist" conception of psychological normality. Normality is viewed as an approximation to an ideal or combination of ideals, such as health, rationality, honesty, integration, maturity, or morality. This literature has justifiably criticized the traditional approaches to this problem, i.e., a "symptom-free" conception, specified as absence of pathological stigmata, or a "centrist" conception, specified either as the statistical average of normally distributed characteristics or as the balance between tendencies, which, in their extremes, are considered pathological, like introversion and extroversion. This paper will present a critique of normative perspectives in personality theory, focusing on the "idealist" position (criticism of traditional approaches may be found in works by Hartman, 1939; Maslow, 1954; Mowrer, 1954; Shoben, 1957, and others), and propose an alternative to these approaches.

A common starting point among authors proposing an idealist conception of normality is the recognition of the enormous extrabiological potential for variation inherent in human beings as evidenced by complex cultural activity, symbolization, and altruism. They stress that these inherently human qualities deserve more serious attention than they are getting from students of personality and should be embodied in the concept of normality. Shoben (1957, p. 185), for example, states, "the fundamental contention advanced here is that behavior is 'positive' or 'integrative' to the extent that it reflects the unique

attributes of the human animal." As the "unique attributes" of humans are extremely varied and frequently repugnant, the writers go on to specify what they mean. Maslow (1954, Ch. 12) lists 17 characteristics of self-actualizing people including acceptance of self, others, and nature; spontaneity; problem centering; the quality of detachment; autonomy; democratic character structure; creativeness; and resistance to enculturation. Bond (1952) mentions freedom to focus energy on main purposes, ability to work and love with ease, and to achieve happiness and efficiency somewhat in proportion to circumstances. Shoben (1957) writes of self-control, personal responsibility, social responsibility, democratic social interests, and ideals. McLaughlin (1950) enumerates emotional independence and self-reliance; balance between giving and getting; relative freedom from egotism, inferiority feelings, and excessive competitiveness; conscience; genital sexuality; constructive aggressiveness; solid sense of reality; flexibility; and adaptability.

Several major difficulties and limitations are inherent in this approach. First, it establishes as absolute desiderata patterns of behavior and values that grow out of a particular cultural context during a particular period of its history. It assumes that what is good for us is good for everybody everywhere. Several authors explicitly recognize that these criteria of normality involve assumptions in the nature of value judgments. Other writers do not, but cultural anthropology and history provide evidence that freedom from inferiority feelings or genital sexuality or constructive aggressiveness, for example, are not and have not been universally considered desirable goals for human behavior.

¹ Part of this paper was prepared while the writer was at the Army Medical Service School, Ft. Sam Houston, Texas.

Second, this approach generally does not take into account the circumstances under which the personality is to function. It implies that variations in circumstances do not determine or even influence the functioning of the personality and precludes the study of how behavior and circumstance covary. Third, it largely assumes that postulated criteria of normality are positively correlated, or can be positively correlated. As to whether they are correlated, reference can be made to several attempts to study "normal, healthy individuals." Maslow (1954) and Bond (1952), for example, went to great efforts to obtain suitable examples. Bond guarded against his own selection biases by studying almost complete populations of student council members and concluded nonetheless that 57% could benefit from ideal mental health services. Maslow grudgingly devotes two pages to the imperfections of self-actualizing people. Roe's (1950) studies of eminent scientists (to be sure, using criteria of success and not of normality) amply indicate that nonnormal characteristics regularly accompany achievement, creativity, and success. Redlich (1952) reviews several studies of "normal" populations in which "rich pathological" material was found.

Perhaps even more significant is the question of whether it is reasonable to anticipate that characteristics of normality such as those under discussion can be positively correlated. Consider, for example, such positively valued attributes as love and independence. Ideal patterns of loving devotion and enduring commitment to one's beloved, coupled with independence, assertiveness, and freedom from infantile demands of succorance, are generally accepted as positive characteristics of the normal; but the conflictual nature of multiple human needs and aspirations is obscured by an approach which does not recognize that love and commitment *necessarily* and inevitably restrict independence.

Finally, a value based conception of normality tends arbitrarily to slant research so as to preclude investigations of the relationships between devalued patterns of behavior and ideal patterns of behavior.

These criticisms are not to be construed as arguments for banning values from considera-

tion in personality theory. The question is where and how to deal with them in a way that is scientifically appropriate. Values and value systems may be studied objectively as variables, and the task of psychology is to assess their relationships with other variables and with behavior. Such an approach does not imply that every moral and ethical system is as good or as workable as any other nor that the search for the "good" in either individual or culture should be minimized, but it does indicate a way in which such problems may be considered at levels of discourse which are germane to science. In this manner science is not faced with questions it cannot answer.

It is significant that the proponents of the point of view under discussion seem to be all for the good life but are rather vague as to what they mean by it, while theologians and ethicists are quite explicit and vigorous defenders of the positions they take. The apparent proclivity for tolerance, understanding, and minimization of differences, characteristic of this school of thought (with few exceptions), conceals and overrides hard differences in the approach to life embodied in different value systems.² Yet, to consider values as variables seems definitely uncongenial to those who hold that the values they profess are the only values worthy of consideration and study. For example, Shoben (1957), while arguing that a normative approach based on a statistical average precludes the criticism of a culture, confuses a conflict of values with a scientific controversy when he disdains the possibility that the storm trooper would be considered the prototype of integrative adjustment in the Nazi culture. Disapproval of Nazi culture has nothing to do with understanding how a man raised in Nazi Germany might be expected to become a storm trooper. Science can do no more than predict the consequences of a value system. Political, religious, or cultural conflicts are fundamentally not resolved in the laboratory or on the couch.

Writers adopting an idealist approach con-

² See Herberg (1957) for a critical discussion of the ethical inadequacies of theoretical orientations in psychology which minimize differences in value alternatives and preclude the conceptualization of the realities of moral conflict in the individual.

cerning psychological normality take a position that is steeped in contradiction. On the one hand, their conception of normality has an absolutist quality that is laced with culture-bound value assumptions. On the other hand, they show marked reservations concerning a conception of normality based on social conformity and they abhor any taint of absolutism. Maslow (1954), for example, specifies autonomy, independence of culture and environment, and resistance to enculturation as characteristics of the self-actualizing (normal) person. In doing so, Maslow evidently means to divorce himself from purely statistical conceptions of social conformity. Certainly he does not mean compulsive autonomy, or compulsive resistance to enculturation, yet his whole perspective, shared with many others, including Fromm (1941), Mowrer (1954), and Shoben (1957), is one of meeting a specified ideal, and this inevitably involves some form of cultural bias, some form of limitation, some constriction of freedom.

In a discussion of this contradiction, Knight (1946) deals with the problem by (a) specifying that freedom is largely a subjective matter; (b) showing that many subjective feelings of freedom are spurious, and (c) concluding that genuine freedom and mental health involve the sense of inner compulsion and conformity based on a rational appreciation by the individual of the assumptions, conditions, motives, and values in his life. It is this bedrock of rationality to which all of these writers ultimately arrive.

Kubie (1954) takes perhaps the extreme position and asserts that the sole criterion for normality is the predominance of conscious and preconscious over unconscious factors in the determination of an act. Since Kubie rules out consciousness, *per se*, as a criterion (a delusion is conscious but not normal), he can only be interpreted to mean that rational, reasonable, reality-oriented behavior is normal. On the surface, this seems to be a satisfactory solution to the problem. It posits a criterion which is neutral as far as value is concerned and can conceivably lead to objective assessment procedures. However, such a point of view seemingly would label all non-rational behavior and motivation as abnormal and in doing so would devalue and im-

plicitly condemn any manifestation of emotional spontaneity. Arguing, as Kubie does, that it is the origin of the act and not its subsequent automatic execution that must be assessed, would nonetheless require, at some point, meticulous evaluation of matters of taste, preferences in art, religious values and beliefs, and love objects in order to fulfill this criterion. Kubie's disclaimer (1954, p. 187), "Here is no unreal fantasy of a 'normal' individual out of whom all the salty seasoning of secondary unconscious motivations has been dissolved," when followed in the next paragraph by "What was unconscious . . . must become accessible enough to self-inspection to become conscious when needed," suggests that his conception of "salty seasoning" is not so spicy after all.

When the search for the normal leads to the elevation of the rational, a host of problems concerning the nature of reason, logic, and rationality may be overlooked. These include the distinctions between a rational process and rational content (Hartman, 1956; Reider, 1950), the illogical and prelogical factors which contribute to the development of intelligence and reason (Piaget, 1953), the role of unconscious and regressive factors in creativity (Kris, 1952), and the process of change in form and content of knowledge and logic through a history which has not ended and whose future course is unpredictable by any reliable means.

Finally, and this criticism is germane to practically every attempt to define normality, it assumes that rational, conscious, and preconscious processes, once established, carry on of their own accord without environmental support and stimulation. Recent studies on the effects of sensory deprivation (Hebb, 1955), the classic studies of perception under distorted conditions (Lawrence, 1949), and many studies of behavior under stress (Lazarus, Deese, & Osler, 1952) indicate that this is simply not the case. To specify consciousness or rationality as the essence of normality merely begs the question.

A truly objective, scientific solution to the problem of defining normality when one is concerned with something more than averaging existing traits and behaviors, when one seeks to embody in this definition a particu-

lar conception of what man may be, is impossible to attain. As soon as values, symbols, altruism, and the like are specified in the definition of normality, objectivity yields to time-bound and culture-bound assumptions which are not subject to scientific verification and which are of limited generality. That each person deliberately or unconsciously has a scheme of values, beliefs, and ethics which imparts to him his uniquely human character, that, indeed, to be a human requires such a scheme, is a proposition that does not establish a *particular* system of values or logic as an absolute criterion of normality.

Psychology should stop asking the question "What is normal?" Personality theory requires another orientation. It seems to me that, for purposes of scientific theory construction and also for practical clinical purposes, a rather different and potentially more fruitful approach can be taken. This would entail the viewpoint that human beings have a variety of potentialities, and that the achievement of certain potentialities may entail certain limitations and that achievement and limitation vary with conditions.

Consider it is the distant, withdrawn personality who is unable to form close relationships who, it has been said, stands the best chance of enduring the stress of a prolonged flight into space. The suspicious mistrusting individual, difficult to get along with, may, because of these very characteristics, be an excellent tax inspector, research scientist, or counterintelligence agent. The vain narcissistic personality can become a great entertainer. Reverse the point. Is great acting possible without narcissism? Is creative research possible without mistrust? Is prolonged isolation tolerable for a gregarious individual? The legitimate task of science is not to stand back smugly and label the various distributions of energy and effort as abnormal or normal. The task of theoretical psychology is to provide the concepts and principles which would enable us to comprehend achievements and failings *and their interrelationships*.

Such an approach focuses on man's behavior as the data available for scientific study. Existing terms and concepts such as repression, motive, conformity, defense mechanism, and habit need not be discarded as

conceptual tools, but with this perspective the scientist does not imply damnation or praise when he uses them to characterize a psychological process. This approach sets no limits on the sophistication of the theories developed to encompass the facts of behavior, but it lessens the likelihood of confounding facts and theories and value assumptions. It avoids positing criterion value systems but freely admits the incorporation of value systems as variables contributing to or limiting different potentialities. It sets the stage for research into the development and exercise of man's potentialities for what, on other grounds, may be considered good or evil, but does not prejudice or condemn. It permits the description and understanding of man as he was, is, and can be.

Several features of the present position have been elaborated in the literature. Marmor and Pumpian-Mindlin (1950) and Jahoda (1953), for example, have stressed the interaction of personality and situational variables in the determination of behavior. Hartman (1951) points out that conflicts are inherent to human existence and (1939) that rationality and freedom do not necessarily imply health, while regression and defense are not necessarily maladaptive. Schafer (1954, p. 172) writes, "The growth and organization of personality appears to require that certain of each individual's potentialities be cultivated and others neglected or even plowed under." Marzolf (1947) concludes, "In fact there is no need for the term 'mental disease.'" Several writers (Darrah, 1939; Jones, 1942; Wile, 1940) are doubtful about the existence of "normal" personalities. But in one way or another, they all (with the exception of Wile, 1940) cling to the traditional perspective and express or reflect some concept of normality.

If the framework proposed is theoretically acceptable, objections may nevertheless be raised on practical grounds. If no valid conception of normality is possible, is treatment ever indicated and, if so, toward what goals is treatment directed? At present, certain personality patterns can be labeled abnormal and treatment is presumably directed toward changing abnormal personalities or patterns of behavior into normal ones. The definition of what is normal is considered critical to

clinical practice and its elimination apparently leaves an enormous conceptual void.

Actually, there are several unwarranted assumptions implicit in the view that a conception of normality is central to clinical practice. First is the assumption that treatment is initiated with the finding of abnormality, however defined. Operationally, treatment is initiated when motivation from some source exists for treatment and when therapists are motivated and prepared to undertake treatment. There is a reciprocal relationship between professional personnel and patient pool in that the professional proffers the expectation of relief from distress and contributes to the development of this anticipation in potential clients. Absence of some relevant form of this expectation seems to be one of the factors underlying failure in the treatment of lower-class patients by predominantly middle and upper-class professionals (Hollingshead & Redlich, 1958). The point is that the initiation of treatment results less from a finding of abnormality than from the meshing of reciprocally intertwined (more or less) motivations.

Second, the notion that the aim of treatment consists of changing the patient from abnormal to normal is unjustified. In many cases, in those perhaps most amenable to current forms of treatment, therapy consists largely of helping the patient to resolve conflicting trends within his personality rather than making him or helping him to become normal. The patient who has successfully completed this type of treatment (in which the therapist can remain relatively neutral with regard to values) does not by any flight of the imagination meet the usual criteria of normality. (The individual who speaks of his having been analyzed or treated psychotherapeutically and implies that he is "normal" is usually labeled a prig.) Furthermore, as already noted, if the patient has been enabled to deal more successfully with certain kinds of problems and conflicts, this may be at the expense of other potentialities.

A third questionable point is the assumption that the goals of treatment are determined by scientific criteria. In many types of cases the resolution of conflict is only a minor part of therapy. Much more than the bolstering or the elimination of defensive patterns

within the personality seems indicated in certain types of schizophrenia, in severely schizoid personalities or in antisocial personalities. Very many of these cases require that a way of life be taught. In other words, certain cases require far more than reorganization of the personality; they require education itself and the inculcation of values. Once again, it seems to me that science can contribute to an understanding of the educative process but that the content of this education is largely a matter that must be decided on other grounds. Here, the articulation and specification of the values under which or toward which the therapist operates is of great importance, *but these choices are not scientific matters*. We do the best we can much like any parent living at a particular time and in a particular place.

This point, that values, symbols, and the like *are* involved in psychotherapy (Hacker, 1945), brings the discussion full circle. Davis (1938, p. 65) in a little cited analysis of the mental hygiene movement has argued that the social function of the mental hygienist "is not that of a scientist but that of a practising moralist in a scientific, mobile world." It seems to me that what Davis is saying is that professionals dealing with "ways of life" inevitably have to rely on *some* consistent framework to fill gaps and resolve conflicts where they exist and that this framework willy-nilly will be that prevalent in the cultural group to which the professionals belong. If this issue may be obscured in the treatment of adults, it is blatantly apparent and quite inevitable in the treatment of children and in the advice given to parents about techniques of child rearing (Erikson, 1950). As a consequence, it seems likely that psychological treatment is a conservative force in any society no matter how radical the theories about psychotherapy sometimes appear to be. But the implicit or explicit commitment to values involved in the practice of psychotherapy is an issue separate and distinct from a scientifically useful specification of the concept of normality. The future development of a scientific understanding of personality depends on a less restrictive perspective which, ironically, should benefit clinical practice as well.

A shift in perspective in personality theory from one which looks upon man as normal or

abnormal or somewhere in between to one which views him as having varying potentialities and limitations under varying conditions, seems to offer an orientation more viable, secure, and fruitful than that prevailing at present. The viable aspect of this position comes out of its capacity to conceive of man in different times and different circumstances. The security features result from the elimination of the necessity for the clinician to rationalize about his own normality (Reider, 1950), permitting his work to be evaluated objectively in terms of its results. Finally, the fruitfulness of this approach rests on its reluctance to condemn, to label abnormal, and its emphasis on the potentialities of every person under the proper conditions. It sets the stage for more extensive investigations into the nature of these potentialities and the conditions of their achievement, all the while keeping before us the pitfalls and limitations that such achievement may involve.

REFERENCES

- BOND, E. D. The student council study: An approach to the normal. *Amer. J. Psychiat.*, 1952, 109, 11-16.
- DARRAH, L. W. The difficulties of being "normal." *J. nerv. ment. Dis.*, 1939, 90, 730-739.
- DAVIS, K. Mental hygiene and the class structure. *Psychiatry*, 1938, 1, 55-65.
- ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.
- FROMM, E. *Escape from freedom*. New York: Rinehart, 1941.
- HACKER, F. J. The concept of normality and its practical significance. *Amer. J. Orthopsychiat.*, 1945, 15, 47-64.
- HARTMAN, H. Psychoanalysis and the concept of health. *Int. J. Psycho-Anal.*, 1939, 20, 308-321.
- HARTMAN, H. Ego psychology and the problem of adaptation. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951.
- HARTMAN, H. Notes on the reality principle. *Psychoanal. Stud. Child*, 1956, 11, 31-53.
- HEBB, D. O. The mammal and his environment. *Amer. J. Psychiat.*, 1955, 111, 826-831.
- HERBERG, W. Freud, religion and social reality. *Commentary*, 1957, 23, 277-284.
- HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.
- JAHODA, MARIE. The meaning of psychological health. *Soc. Casewk.*, 1953, 34, 349-354.
- JONES, E. The concept of the normal mind. *Int. J. Psycho-Anal.*, 1942, 23, 1-8.
- KNIGHT, R. P. Determinism, "freedom" and psychotherapy. *Psychiatry*, 1946, 9, 251-262.
- KRIS, E. *Psychoanalytic explorations in art*. New York: International Univer. Press, 1952.
- KUBIE, S. The fundamental nature of the distinction between normality and neurosis. *Psychoanal. Quart.*, 1954, 23, 167-204.
- LAWRENCE, M. *Studies in human behavior*. Princeton: Princeton Univer. Press, 1949.
- LAZARUS, R. S., DEESE, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, 49, 293-317.
- McLAUGHLIN, J. T. Normality and psychosomatic illness. *Ment. Hyg.*, N. Y., 1950, 34, 19-33.
- MARMOR, J., & PUMPIAN-MINDLIN, E. Toward an integrative conception of mental disorder. *J. nerv. ment. Dis.*, 1950, 111, 19-29.
- MARZOLF, S. S. The disease concept in psychology. *Psychol. Rev.*, 1947, 54, 211-221.
- MASLOW, A. H. *Motivation and personality*. New York: Harper, 1954.
- MOWRER, O. H. What is normal behavior? In L. A. Pennington & I. A. Berg (Eds.), *An introduction to clinical psychology*. (2nd ed.) New York: Ronald, 1954. Pp. 58-88.
- PIAGET, J. *The origin of intelligence in the child*. London: Routledge & Kegan Paul, 1953.
- REDLICH, F. C. The concept of normality. *Amer. J. Psychother.*, 1952, 6, 551-569.
- REIDER, N. The concept of normality. *Psychoanal. Quart.*, 1950, 19, 43-51.
- ROE, ANNE. The use of clinical diagnostic techniques in research with normals. In M. Reymert (Ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950. Pp. 336-342.
- SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954.
- SHOBEN, E. J. Toward a concept of the normal personality. *Amer. Psychologist*, 1957, 12, 183-189.
- WILE, I. S. What constitutes abnormality? *Amer. J. Orthopsychiat.*, 1940, 10, 216-228.

(Received March 31, 1959)

AGE DIFFERENCES IN THE INTERSCALE FACTOR STRUCTURE OF THE GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY

A. W. BENDIG

University of Pittsburgh

One of the relatively unexplored areas in objective personality measurement concerns age differences in test scores. Many manuals for personality tests state that an age correction must be made in the raw scores before the scores can be referred to the test norms, or norms are given for different age groups without any apparent interest in what these age differences mean in terms of possible changes in personality integration. A second question relates to possible age differences in the internal structure or interrelationships among the scales included in the personality inventory. Most of these inventories have been developed, by factor analysis or some other statistical technique, on relatively young groups of subjects (the traditional college sophomore being widely used), and the scale scores are subsequently interpreted under the assumption that each scale means the same thing or measures the same underlying personality dimension at all age levels. This appears to be a hazardous assumption in the absence of data on the stability of personality test scores over the mature age range.

A personality inventory that is widely used with nonpsychiatric populations is the Guilford-Zimmerman Temperament Survey (Guilford & Zimmerman, 1949) which contains 10 scales of General Activity (G), Restraint (R), Ascendancy (A), Sociability (S), Emotional Stability (E), Objectivity (O), Friendliness (F), Thoughtfulness (T), Personal Relations (P), and Masculinity (M). These scales are revised and condensed versions of 13 scales that had been developed by factor analyses of test items and published in previous inventories (Guilford, 1940; Guilford & Martin, 1943a, 1943b). Although the scales presumably measure 10 orthogonal or independent

first-order personality dimensions, scores from certain scales typically show rather large intercorrelations. For example, correlations between the A and S scales usually are about .68 and between the E and O scales average around .66. However, the correlations among other scales are quite low. This suggests that scores from these scales may be organized at a higher level of personality into a smaller number of more general factors describing broader syndromes of behavior. Factor analyses of the correlations among the GZTS scales would provide evidence as to the number of such factors involved and an empirical method of describing these syndromes. However, as pointed out by Guilford and Zimmerman (1956, p. 2), these second-order or "super" factors represent factors common to the inventory scores and do not necessarily represent the second-order organization among the original first-order inventory factors which the scores were designed to measure.

The research reported below was designed to provide evidence as to the factor organization of the GZTS scores and to investigate possible age differences in both the average scale scores and the factor loadings of the individual scales.

PROCEDURE

Subjects. The Ss were 400 men whose GZTS scores were available from the files of a local psychological service agency.¹ Over 800 male and female Ss were tested during the year 1953, and the cases were randomly selected from this larger group with the restrictions on randomness being: (a) only male Ss were included; (b) 100 Ss were selected from each of the following age (at time of testing) ranges—20-29, 30-39, 40-49, and 50-59 years; and (c) the

¹ The author is indebted to James Horgan and the research staff of Psychological Service of Pittsburgh for their kindness in making these data available.

TABLE 1

SUMMARY OF ANALYSES OF VARIANCE OF AGE DIFFERENCES IN MEAN SCORES OF MEN ($N = 400$)
ON THE GUILFORD-ZIMMERMAN TEMPERAMENT SURVEY

GZTS Scale	Age Groups								Error Mean Sq. (<i>df</i> = 396)	Significance of Age Differences (<i>F</i>)			
	20-29		30-39		40-49		50-59			Groups (<i>df</i> = 3)	Linear (<i>df</i> = 1)	Quad- ratic (<i>df</i> = 1)	Cubic (<i>df</i> = 1)
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>					
G	18.86	4.69	18.87	5.06	18.85	5.17	17.38	5.16	25.49	2.15	3.90*	2.15	.40
R	18.87	4.21	20.18	4.32	20.64	3.52	20.66	4.00	16.36	4.33**	10.39**	2.54	.05
A	19.83	4.97	18.86	5.00	17.61	5.63	16.74	5.65	28.62	6.47**	19.33**	.01	.76
S	23.77	5.01	22.88	4.76	21.84	5.12	20.60	6.22	28.45	6.56**	19.56**	.11	.00
E	21.87	4.75	21.57	5.25	20.96	4.90	21.58	4.27	23.33	.63	.47	.91	.51
O	21.87	4.75	21.57	5.25	20.96	4.90	21.58	4.27	23.33	1.87	3.32	.60	1.69
F	17.39	4.73	18.42	4.81	17.45	4.75	17.99	4.71	22.79	1.04	.15	.26	2.70
T	19.63	4.28	19.67	4.56	20.53	3.74	20.31	4.01	17.47	1.18	2.41	.10	1.03
P	21.61	5.38	23.53	3.90	22.11	4.41	22.77	4.10	20.31	3.41*	1.04	1.95	7.23**
M	20.96	3.20	20.14	3.55	19.54	3.25	19.14	3.51	11.55	5.43**	15.90**	.38	.00

* $P < .05$.

** $P < .01$.

age distribution within each age decade was kept as rectangular as possible. The mean ages for the four groups of Ss were 24.8, 34.5, 44.2, and 54.0, while the standard deviations of ages within each group were 4.4, 2.9, 3.0, and 2.5. The Ss had been referred to the agency by local business firms for psychological testing for employment, promotion, or company personnel survey purposes and had been given a large battery of aptitude, interest, and temperament tests including the GZTS.

Method. The raw scores from each of the 10 GZTS scales were subjected to a one-criterion (between-within) analysis of variance with the age of the Ss (four decades) being the independent variable. The between-age groups sums of squares were further analyzed into linear, quadratic, and cubic components by the method of orthogonal polynomials (Grant, 1956) and tested for significance. The F ratio test of the between-groups sums of squares tests for the presence of significant differences among the GZTS score means of the age groups, but does not specify the form of the functional relationship between age and the scale means. The further breakdown of the sums of squares tests whether the relationship is a straight line (linear), simple curvilinear (quadratic), or more complex (cubic) change in means with increasing age.

The scores within each age group were intercorrelated by the usual product-moment method, and each of the four matrices of correlations was factored by the complete centroid method, with five factors being extracted from each matrix. Each successive factor was tested for significance using Humphrey's rule (Fruchter, 1954, pp. 79-80), and the apparently significant factors were rotated to orthogonal simple structure using the normalized varimax criterion (Kaiser, 1958).²

² The author's appreciation is extended to Gary Lotto and William B. Kehl of the University of Pittsburgh Computation and Data Processing Center for making their facilities available for this phase of the statistical analysis.

RESULTS

A summary of the analyses of variance of the GZTS scales can be found in Table 1, which gives the means and standard deviations of the age groups on each scale, the within-age groups (error) mean square from the analyses of variance, and the F ratio tests of significance of the between-age groups mean square and its linear, quadratic, and cubic components. The age differences in mean scores were significant at the .01 level for four scales (R, A, S, and M), and one additional scale (P) was significant at the .05 level. The linear trend in the means was significant at the .01 level for four scales (R, A, S, and M) and at the .05 level for the G scale. The cubic trend was significant at the .01 level for the P scale. None of the quadratic components was significant even at the .10 level. The means of the A, S, and M scales show a linear (straight line) decline with increasing age, while the G scale is essentially level until age 50 and shows a sharp drop between the 40-49 and 50-59 age groups. The R scale mean scores increase with age with some deceleration in the rate of increase. The P scale means tend to increase with age, but the increase is irregular, with the largest P scale mean found for the 30-39 age group. Scales E, O, F, and T show no systematic variation with age.

The first two centroid factors were significant at the .01 or .05 levels within all four age groups, while the third factor was sig-

nificant at the .01 level for the 20-29 and 30-39 groups but was not significant for the two older groups of Ss. Neither the fourth nor the fifth factor approached significance at the .05 level within any age group. For comparability between age groups, the first three factors were retained in the subsequent normalized varimax rotations with the remaining two factors being dropped. The percentages of variance accounted for by the first three orthogonal factors within each group were 54.1, 51.3, 45.1, and 44.3.

The rotated factor loadings and communalities for each GZTS scale can be found in Table 2. The approximation to simple structure was moderately successful with slightly over half of the loadings having absolute values less than .20. The first factor is characterized by consistently high loadings on the O, P, and F scales for all age groups and high to moderate loadings on the E and M scales only for the younger groups. The second factor shows consistently high loadings on Scales S and A and moderate loadings on Scale G for all age groups and gradually increasing loadings on Scale E as the age of the Ss increases. The third factor loads Scales R and T only with no consistent loadings on any other scale. The O scale has a small but consistent loading on the second factor, and Scales T and R have slight negative (T) and positive

(R) loadings on the first factor. The similarity between comparable factors appearing in the four analyses was quantified by the method suggested by Wrigley and Neuhaus (1955). The indices of factor similarity for the first rotated factor ranged from .98 to .83, with the median of the six indices being .93. Indices for the second factor ranged from .96 to .68 with the median also being .93, while the range for the third factor was .96 to .41 with a median of .85. The indices were consistently lowest when factors were compared between the two groups at the extremes of the age continuum, 20-29 and 50-59, and is probably attributable to the changing factor composition of Scales E and M.³

DISCUSSION

The first factor found in the factor analyses of the GZTS scales included the scales of Objectivity, Friendliness, and Personal Relations. These three scales are GZTS versions of the Objectivity, Agreeableness, and Cooperativeness scales developed by Guilford and Martin (1943b) from a factor analysis of items

³ Tables giving the original correlation matrices, unrotated factor loadings, transformation matrices, tests of the significance of the factors, and indices of factor similarity have been deposited with the American Documentation Institute. Order Document No. 6120, remitting \$1.25 for microfilm or \$1.25 for photocopies.

TABLE 2
ROTATED ORTHOGONAL FACTOR LOADINGS (DECIMAL POINTS OMITTED) OF THE GUILFORD-ZIMMERMAN
TEMPERAMENT SURVEY SCALES FOR FOUR AGE GROUPS OF MEN
(N = 100)

GZTS Scale	Age 20-29				Age 30-39				Age 40-49				Age 50-59			
	Fr	SA	EI	h ²	Fr	SA	EI	h ²	Fr	SA	EI	h ²	Fr	SA	EI	h ²
G	-.03	.54	.06	.30	.11	.55	-.20	.36	-.09	.38	-.13	.17	-.07	.43	.00	.19
R	.22	.03	.68	.52	.11	-.07	.65	.44	-.07	-.11	.37	.15	.31	-.03	.52	.36
A	.16	.77	.20	.67	.13	.84	.09	.73	.20	.82	.16	.75	.12	.73	-.07	.55
S	.10	.85	-.11	.74	.11	.73	.00	.54	.10	.85	.08	.74	.17	.69	-.32	.60
E	.77	.32	-.06	.70	.70	.48	.29	.81	.37	.46	-.32	.45	.35	.70	.15	.64
O	.90	.19	-.02	.84	.78	.34	-.01	.72	.73	.19	-.41	.75	.72	.30	-.04	.62
F	.57	-.12	.13	.36	.64	-.14	.34	.55	.55	-.34	-.31	.51	.77	-.02	.15	.61
T	-.18	.09	.69	.51	.34	.05	.49	.35	-.12	.19	.56	.36	-.06	-.04	.44	.20
P	.62	.26	-.04	.46	.51	.20	-.12	.31	.74	.02	-.05	.55	.74	.08	.03	.55
M	.53	-.13	.02	.30	.53	.10	-.16	.32	.34	.07	-.02	.12	.22	.22	-.14	.12

Percent- age of variance	25.1	18.5	10.2	53.8	22.2	19.6	9.5	51.3	17.1	19.6	8.7	45.4	19.8	18.3	6.4	44.5
--------------------------------	------	------	------	------	------	------	-----	------	------	------	-----	------	------	------	-----	------

culine groups of younger male Ss, individual differences in Masculinity are positively associated with differences in Friendliness, while within the generally more feminine groups of older Ss the correlation between Masculinity and Friendliness disappears.

All three of the scales loading on the factor of Social Activity show decreasing mean scores with increasing age. The loading of the GZTS G (General Activity) scale on factor SA slightly decreases with age with the size of the loading tending to parallel the size of the G scale mean. The older male Ss appear on the GZTS to withdraw more from vigorous social activities and to be more submissive in the social world. This does not imply a withdrawal from society but, rather, a reorientation of their role in society toward a less competitive position that is more free of the demands of social leadership.

The factor of Extraversion-Introversion shows some changes with age, primarily in the direction of greater introversion among older Ss. This trend was statistically significant only for the Restraint component of EI, with Thoughtfulness not increasing as markedly among the older groups. Scores on the two scales loaded for the EI factor increase up to the 40-49 age group, and no change in mean R and T scores is apparent between the 40-49 and 50-59 age groups.

Obviously this cross sectional study does not imply that male Ss show personality changes with age that are discussed above, but that these personality differences existed among groups of Ss who were born in the years (approximately) 1928, 1918, 1909, and 1899. Whether the trait differences apparent in 1953 reflect age changes in the Ss or differences in the environmental factors influencing personality formation in childhood during these early years cannot be assessed from this type of study.

SUMMARY

Scores on the 10 Guilford-Zimmerman Temperament Survey scales were obtained for four groups of male Ss ($N = 100$) in different age decades: 20-29, 30-39, 40-49, and 50-59 years. Scores on the General Activity, Ascendancy, Sociability, and Masculinity scales significantly and linearly decreased with increasing age, while the Restraint scores linearly

increased with age and scores from the Personal Relations scale showed an irregular increase. Factor analyses of the scale intercorrelations resulted in three major orthogonal factors: Friendliness (Fr), Social Activity (SA), and Extraversion-Introversion (EI). Scales of Objectivity, Friendliness, and Personal Relations consistently loaded on the Fr "super" factor, Ascendancy and Sociability loaded on the SA factor, while Thoughtfulness and Restraint loaded for the EI factor. The loading of Masculinity on the Fr factor showed a progressive decline with increasing age. Emotional Stability inconsistently loaded on both Fr and SA factors and it was suggested that the E scale may represent a fourth second-order factor (Emotionality) present in the GZTS inventory.

REFERENCES

- EYSENCK, H. J. *The structure of human personality*. London: Methuen, 1953.
- FRUCHTER, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.
- GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, 53, 141-154.
- GUILFORD, J. P. *An inventory of factors STDCR*. Beverly Hills, Calif.: Sheridan, 1940.
- GUILFORD, J. P., & MARTIN, H. G. *An inventory of factors GAMIN*. Beverly Hills, Calif.: Sheridan, 1943. (a)
- GUILFORD, J. P., & MARTIN, H. G. *Personal inventory: Manual of directions and norms*. Beverly Hills, Calif.: Sheridan, 1943. (b)
- GUILFORD, J. P., & ZIMMERMAN, W. S. *The Guilford-Zimmerman Temperament Survey: Manual of instructions and interpretations*. Beverly Hills, Calif.: Sheridan, 1949.
- GUILFORD, J. P., & ZIMMERMAN, W. S. Fourteen dimensions of temperament. *Psychol. Monogr.*, 1956, 70(10, Whole No. 417).
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.
- KLUGH, H. S., & BENDIG, A. W. The Guilford-Zimmerman Temperament Survey and intelligence. *Psychol. Newsltr.*, NYU, 1959, 10, 96-97.
- LOVELL, C. A study of the factor structure of thirteen personality variables. *Educ. psychol. Measmt.*, 1945, 5, 335-350.
- THURSTONE, L. L. The dimensions of temperament: Analysis of Guilford's thirteen personality scores. *Psychometrika*, 1951, 16, 11-20.
- WRIGLEY, C. F., & NEUHAUS, J. O. The matching of two sets of factors. *Amer. Psychologist*, 1955, 10, 419.

SIZE-DISTANCE JUDGMENT IN ORGANIC MENTAL DEFECTIVES¹

NOËL JENKIN

University of Sydney

AND SALLY A. MORSE

University of Virginia Hospital

Koffka's (1935) discussion of the constancies gave a controversial status to the question concerning progressive development of size constancy in childhood. Of recent years, empirical evidence has accumulated to resolve this controversy. Lambercier (1946a, 1946b), Piaget and Lambercier (1943, 1951, 1956), and Zeigler and Leibowitz (1957) all agree as to the existence of a developmental trend in objective size-distance judgments, which reaches nearconstancy or overconstancy in adulthood.

The present study focuses upon effects of the type of impaired development which is due to organic brain damage. It was expected that organic mental defectives, due to their impaired organizing and integrating abilities, would make judgments which fall short of size constancy, relative to the judgments of normal persons of equivalent age. The defective individuals would thus occupy a place on the developmental curve which is normally marked by a younger age group.

Since an obtained difference in mean size-matches between organic mental defectives and normals could be attributed to the difference in mental age between these groups, it was planned to use a familial mentally retarded group as a control. The familials would have mental ages similar to those of the organics, but would lack the factor of brain damage which the present hypothesis specifies as the cause of judgments atypical for a given

chronological age. Success in judging the physical size of a distant object is considered to be an ability developed through normal visual experience and not primarily dependent on intelligence. It was therefore expected that the familial group would make size-distance judgments which are closer to those of the normal group than to those of the organic group.

A supplementary aim of the investigation was to study the trends of the size-matches when the comparison object was presented at three different distances. Previous work by Jenkin (1957, 1959) has shown that with normal adult Ss, objective size-matches of a standard at a constant distance become progressively reduced as the comparison object recedes from the point of observation. This tendency, which has been interpreted as an increase of apparent size with increasing distance, was thought to be probably a product of visual learning. If this is so, an interaction of groups by distances would be found when the organics are compared with other groups. The expected direction of the relationship was towards organics showing less difference between distances than either normals or familials.

METHOD

Subjects. Four groups of Ss were used, each group numbering 16 cases and including both sexes. Cases of gross visual defect were excluded from each group. The organic group consisted of mentally retarded residents of the Training School at Vineland and the Vineland State School. Each member of this group had been classified in the records of the respective institution as brain damaged from causes occurring prenatally or early in life. The median age of this group was 18 yr. 6 mo., and the median men-

¹ This study was conducted when the authors were at the Training School at Vineland. Thanks are due to the superintendent and staff of the Vineland State School, whose cooperation made possible collection of some of the data.

al age (Binet, or equivalent) was 7 yr. 10 mo. The amillial group was obtained from the same institutions. Its members were classified in the records as mentally retarded cases of familial etiology. The median age was 15 yr. 9 mo., and the median mental age (Binet, or equivalent) was identical with that of the organic group, 7 yr. 10 mo. The normal adolescent group consisted of students from a nearby junior high school who agreed to serve as Ss for monetary compensation. The median age was 14 yr. No mental ages were available, but it was ascertained that the students were all achieving at school at a rate normal for their age. The normal adult group consisted of employees of the Training School at Vineland, most of whom were young professional personnel. The median age was 23 yr.

Apparatus. The apparatus consisted of white cardboard squares. The standard card, 4 in. square, was attached by a steel wire to a ringstand and was placed directly in front of S at a distance of 320 in. from his eyes. The comparison series was a set of squares ranging from 2½ in. to 6 in. by steps of ½ in. They were individually displayed at the same height as the standard upon a platform which was situated on S's right. Concealed pins and a slight depression, invisible to S, served to maintain the comparison cards in vertical position in the middle of the platform. The angular separation between the standard and the comparison was 87°. Homogeneous fields against which both stimuli were viewed were provided by wooden screens painted a dull green. Little daylight entered the room, and the stimuli were illuminated by 75-w. floodlights placed so as to give an illumination reading of 11 foot candles (measured by a Macbeth Illuminometer) from a correctly positioned square.

Design. Each S judged the size of the standard with the comparison object placed at three different distances: 20 in., 80 in., and 160 in. Two trials were conducted at each distance, one an ascending and the other a descending series. Each group of Ss (organic, familial, adolescent, adult) was divided into two subgroups, A and B. Subgroup A experienced the different distances in the order 80 in., 160 in., 20 in.; and Subgroup B experienced the order 20 in., 160 in., 80 in. Half the members of each subgroup began with an ascending series and the other half began with a descending series.

Procedure. The height of the stimulus objects and S's eye level were properly matched. The platform was placed at the distance appropriate for viewing the comparison object on the first trial. S was then instructed about the nature of the judgment required. He was asked to respond to each comparison card with an answer of "smaller," "the same," or "bigger." It was explained that "the same" meant that the comparison card would exactly match the standard in its physical dimensions. The concepts of "smaller" and "bigger" were illustrated with demonstration cards, so chosen as to lie outside the range of comparison cards used in the experiment. The concept of "equality" was illustrated with a pair of cards 3 in. square, held in E's hands, then superim-

posed and rotated before S's eyes. Two practice trials were then conducted, one ascending and the other descending. A pilot study had indicated that, with retarded Ss, at least this much experience was necessary for the thresholds to attain reasonable stability. Data from the practice trials were therefore excluded from the analysis. S was required to look first at the standard and then at the comparison, keeping his head erect and in the same position for each judgment. He was not prevented from repeatedly looking back and forth if he wished to do so. Within predetermined limits, E varied the size of the card with which each series began, and also randomly switched from one set of comparison cards to a second identical set.

RESULTS

Table 1 shows the mean size-matches at each of the three distances for the four groups of Ss. It will be seen that the grand mean for the organic group is less than 4 in., which latter value would represent perfect constancy. The grand mean for every nonorganic group is higher than 4 in.

The results of an analysis of variance of the data are shown in Table 2. Variability among the four groups is statistically significant ($P < .01$). The three distances also represent a significant source of variance. The specific hypotheses of the experiment are answered by *t* tests of the significance of the difference between pairs of groups. With 30 *df*, the value obtained for the organics vs. familials comparison was $t = 2.32$, $p < .05$. A test of the organics vs. adolescents comparison yielded $t = 2.63$, $p < .02$, and between the organics and adults, *t* reached a value of 4.06, $p < .001$. All comparisons of nonorganic pairs of groups, as expected,

TABLE 1

MEAN SIZE-MATCHES, IN INCHES, OF A 4-IN. STANDARD^a
BY 4 GROUPS OF SUBJECTS

Group	Distance of Comparison Object			
	20 In.	80 In.	160 In.	Grand Mean
Organic defectives	3.777	3.805	3.746	3.776
Familial defectives	4.178	4.250	3.982	4.137
Normal adolescents	4.246	4.195	4.113	4.185
Normal adults	4.605	4.355	4.262	4.407

^a The standard was 320 in. from the point of observation.

TABLE 2
ANALYSIS OF VARIANCE OF SIZE-MATCHES

Source	df	MS	F
Groups (G)	3	210.25	5.66*
Order (O)	1	103.55	2.79
G \times O	3	49.15	1.32
Ss	56	37.16	
Distances (D)	2	33.57	7.81**
D \times G	6	7.96	1.85
D \times O	2	5.51	1.28
D \times G \times O	6	3.57	0.83
Residual	112	4.30	

* $P < .01$.

** $P < .001$.

showed nonsignificant differences, p being greater than .05 in every case.

It is noteworthy that the organic group made a mean size-match significantly smaller than that of the familial group, while the size-match of the latter group did not differ significantly from those of either of the normal groups. The atypical performance of the organics cannot be attributed to low measured intelligence, since the two retarded groups were equated on this variable.²

The effort to match the ages of the organics, familials, and normal adolescents was not entirely successful. The unavoidable error of having an organic group with too high a median age is, however, in a direction which weakens the chances of a positive finding. Both the familial and the normal adolescent groups are, on the average, younger than the organic group and, on the developmental hypothesis, this would tend to minimize any differences which may be due to organicity. Nevertheless, our hypothesis concerning the relationship of the organics to the other groups is strongly supported by the data.

An examination of Table 1 shows that the trend of mean size-matches among the three different distances of the comparison appears as expected for the normal groups, i.e., towards smaller matches as distance increased. Also as expected, there is less variability be-

tween distances in the case of the organic group. Table 2 indicates, however, that the differences among these trends are not great enough to produce a significant groups \times distances interaction.

DISCUSSION

The present study has demonstrated that organic mental defectives tend to underestimate the size of a distant object. Normal persons tending to be younger than the defectives are inclined to overestimate the size of a distant object, and so are normals who tend to be older than the defectives. The difference between the performances of organics and normals cannot be attributed to the difference in Binet mental age, as shown by the performance of a control group of familial etiology.

These results, which support the main hypothesis of the investigation, may be usefully compared with data from a study of normal children and adults by Piaget and Lambercier (1951). A group of children, aged from 7-8 yr., made a mean objective match which fell short of size constancy. The next group, aged from 8-10 yr., slightly exceeded size constancy, and the three remaining groups respectively gave average values which showed a progressive overestimation as age increased. The direction and extent of the difference between Piaget and Lambercier's adolescent (12-14 yr.) and adult groups resemble what was found from the comparable groups of the present study. Our organic group, however, behaved like the young children (7-8 yr.) of Piaget and Lambercier, even though our cases were very much older.

It appears reasonable, therefore, to interpret the behavior of our organic group as indicating immaturity of size-distance judgment, due to brain damage occurring prenatally or early in life. A firm conclusion of this type is not at present possible, since the present authors did not have the opportunity to study a group of individuals who had suffered recent brain damage. A future experiment might determine whether or not such persons behave like normals or like young children in the size-distance judging situation, and the role of learning would thus be further clarified.

² A further test of the possible role of intelligence was made by computing the rank correlations between mental ages and mean size-matches. Nonsignificant values of $-.22$ for the organic group and $-.05$ for the familial group were obtained.

In the meantime, the authors consider that the present results are generally consistent with other work on perception in organic mental defectives, summarized by Strauss and Kephart (1955) and by Gallagher (1957), and might usefully be interpreted along broadly similar lines. This is to the effect that prenatal or early brain damage impairs the development of integrative cerebral mechanisms which are necessary for normal perceptual functioning (Strauss & Kephart, 1955, pp. 57-58). Organics thus tend to show features such as a fragmentary response to a complex pattern of stimulation, inadequate perceptual discrimination, and a failure of synthesis in a situation where several sources of information are normally combined.

It is the latter feature which in particular seems to have been demonstrated in the present study. It cannot be pretended that a fully satisfactory theory of size constancy exists, but it is known that its manifestation depends upon the apprehension of various depth cues and the integration of these with other sources of information. The proximal stimulus size and also identification and categorization of the object displayed are probably included among these sources. The chronically impaired ability to integrate several sources of information and the consequent weakness of the information which is given by experience are factors probably responsible for the tendency of the organic defective to rely more heavily than other groups upon the proximal stimulus size.

It is of interest to note in this connection that "normal" perception is not necessarily the most veridical perception. Table 1 indicates that the organic group was, on the average, more successful than the adult group in matching the objective size, and the same statement may be made about the young children in Piaget and Lambercier's study. As pointed out by Jenkin and West (1958), in connection with the size-weight illusion, the gross degree of error measured in some perceptual situations is not an index of perceptual inefficiency; rather, it is the opposite. This point of view leads us to interpret the behavior of the organics as tending to be "stimulus bound," whereas the normals, by virtue of their superior organizing ability and

more adequate fund of experience, tended more to respond to the total situation, including its illusory³ aspects.

SUMMARY

Binocular objective size-matches of a standard 4-in. square situated 320 in. from the observer were obtained at three different distances of the comparison object. Under conditions of strong depth cues and ample illumination, judgments were made by four groups of Ss ($N = 16$ per group): (a) familial type mental defectives, (b) brain damaged mental defectives, (c) normal adolescents, (d) normal adults. As hypothesized, the organics made matches significantly closer to the visual angle than those of the other groups. There was no significant difference between the familials and the normals. As expected, variation among the different distances was smaller in the case of the organic group, but this trend was not statistically significant. The data were discussed in terms of perceptual learning, and it was suggested that the performance of the brain damaged cases is due to chronically impaired ability to integrate several sources of information.

REFERENCES

- GALLAGHER, J. J. *A comparison of brain injured and non-brain-injured mentally retarded children on several psychological variables*. Lafayette, Ind.: Child Developm. Publ., 1957.
- JENKIN, N. Effects of varied distance on short-range size judgments. *J. exp. Psychol.*, 1957, 54, 327-331.
- JENKIN, N. A relationship between increments of distance and estimates of objective size. *Amer. J. Psychol.*, 1959, 72, 345-363.
- JENKIN, N., & WEST, NOËL. Perception in organic mental defectives: An exploratory study. I. The size-weight illusion. *Train. Sch. Bull.*, 1958, 55, 5-10.
- KOFFKA, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
- LAMBERCIER, M. La constance des grandeurs en comparaisons sériales. *Arch. Psychol., Genève*, 1946, 31, 79-282. (a)
- LAMBERCIER, M. La configuration en profondeur dans la constance des grandeurs. *Arch. Psychol., Genève*, 1946, 31, 287-324. (b)

³ All Ss were shown the standard at close range (20 in.) after completion of the experiment. A large majority of the normal adolescents and adults remarked that it looked smaller than was thought under conditions of the test.

- PIAGET, J., & LAMBERCIER, M. La comparaison visuelle des hauteurs à distances variables dans le plan frontoparallèle. *Arch. Psychol., Genève*, 1943, 29, 175-253.
- PIAGET, J., & LAMBERCIER, M. La comparaison des grandeurs projectives chez l'enfant et chez l'adulte. *Arch. Psychol., Genève*, 1951, 33, 81-130.
- PIAGET, J., & LAMBERCIER, M. Grandeurs projectives et grandeurs réelles avec étalon éloigné. *Arch. Psychol., Genève*, 1956, 35, 257-280.
- STRAUSS, A. A., & KEPHART, N. C. *Psychopathology and education of the brain-injured child*. Vol. II. *Progress in theory and clinic*. New York: Grune & Stratton, 1955.
- ZEIGLER, H. P., & LEIBOWITZ, H. Apparent visual size as a function of distance for children and adults. *Amer. J. Psychol.*, 1957, 70, 106-109.

(Received April 13, 1959)

IMPULSE REPRESSION AND EMOTIONAL ADJUSTMENT

HARRY GRATER¹

Michigan State University Counseling Center

Theories of personality development have emphasized what behavior the individual considers acceptable or right and what behavior he considers unacceptable or wrong. These judgments of rightness or wrongness form important parts of psychological concepts such as value systems, super ego, ego-ideal and self concept. The basic premise is that while very young the child learns that certain types of behavior resulted in parental approval, while other forms of behavior bring disapproval. Since the young child's security and sense of well being are so closely bound to parental evaluation, he soon learns to avoid the disapproved forms of behavior and eventually to repress impulses to such acts. The child experiences anxiety if he behaves in a nonacceptable fashion or even if impulses to such behavior are experienced.

Since the above position places such strong emphasis on conflicts between impulses to behavior and forces acting to inhibit behavior, it could be called an impulse repression theory of anxiety.

The experimental evidence concerning the validity of the impulse repression position is contradictory. Stone and Landis (1953) found that high school seniors who were raised in a democratic family structure experienced fewer conflicts with their families than did those raised in an autocratic family. Watson (1934) studied the effects of permissive vs. strict home discipline and found that subjects raised under strict home discipline reported more indications of emotional conflict; however, Watson (1957) in a later study found no difference in the emotional adjustment of children

raised in strict and permissive families. Brown (1942) found no relationship between the attitudes of parents toward the child's behavior and the child's adjustment. Lasko (1952) reported that children raised by laissez faire parents gave indications of emotional maladjustment.

Two factors limit the effectiveness with which one can use the above empirical data to evaluate the impulse repression theory of anxiety. First, as previously indicated, the above studies are contradictory. Second, they have tended to emphasize the democratic-autocratic continuum. It is quite possible for a family to have an autocratic structure and yet allow the child relatively free expression of certain basic impulses. Thus, determining that an autocratic family structure results in greater maladjustment does not indicate that this maladjustment resulted from the repression of basic impulses.

Since the impulse repression theory already presented depicts anxiety as resulting from conflicts between impulses and repression, the child who has been taught a large number of behavior restrictions by the significant people in his life should experience greater anxiety than the child brought up under more liberal conditions. The above should be true for there would be more opportunity for the child with a high set of behavior standards to experience conflicts between impulses and standards, as well as between behavior and standards. If the child with the high set of behavior standards experiences more conflict, he should have also developed more symptoms of these conflicts, that is, more neurotic symptoms.

This research was designed to explore the following questions which are related to the impulse repression theory of anxiety already presented. Can subjects who report a high number of neurotic symptoms be distinguished

¹ This paper is based upon a doctoral dissertation completed at Pennsylvania State University. The writer should like to acknowledge the advice and suggestions of William U. Snyder under whose direction this study was done.

from subjects who report a smaller number of neurotic symptoms on the following factors:

- (a) The number of moral and ethical standards of behavior held by the subject?
- (b) The frequency with which they engaged in behavior involving a moral or ethical judgment?
- (c) The number of moral and ethical behavior standards held by one of the significant adults in the subject's life as perceived by the subject?
- (d) The number of moral and ethical behavior standards held by the significant adult in the subject's life?
- (e) The frequency with which the subject engaged in behavior which he considered wrong, perceived a significant adult in his life as considering wrong, or is considered wrong by a significant adult in his life.

METHODS AND PROCEDURES

Single, white, Protestant university females were selected as the population to be used in this research. Mothers were selected as the significant adult in the subject's life.

The Minnesota Multiphasic Personality Inventory was chosen as the instrument to measure emotional adjustment. Since the experimental questions referred to neurotic symptoms, these three scales forming the neurotic triad, hypochondriasis (*Hs*), depression (*D*) and hysteria (*Hy*), were selected as measures of neurotic symptomatology.

In order to obtain items for a scale of behavior standards, a questionnaire was sent to 175 college females. They were asked to list the significant moral or ethical decisions that females of their age were called upon to make. A list of 67 items was compiled from the returns to this questionnaire.

These items were used to form three scales. Each item was contained on all scales, but the subjects were asked to respond in three different ways. On the first scale, they indicated whether they believed the behavior described would be morally or ethically right or wrong for all females of their age. On the second scale, they were asked to indicate whether they believed their mother would consider their behavior right or wrong for females of their age, and on the third scale, they were asked to indicate whether or not they had engaged in the behavior described. There were called the Daughter Wrong, Perceived Mother Wrong, and Not Done scales, respectively.

These three scales were administered to a sample of 100 college females. The scales were scored by

counting the number of items checked wrong on the Daughter Wrong and Perceived Mother Wrong scales, and by counting the items marked as not done on the Not Done scales. The results were used in an item analysis. Responses of subjects falling in the top and bottom quarters were compared for each of the three scales. Phi coefficients were calculated graphically for each item. Those items which discriminated between the top and bottom quarters on one or more scales at the 5% level of confidence were retained. A total of 62 items were included on the basis of this criterion.

The items included on the final scale measured behavior or attitudes involving hostility, heterosexuality, religion, college regulations, and some items less easily described which involve some of the social behavior expected of university females. The following are examples of each of the above types of items:

- (a) being critical of your parents
- (b) having sexual relations prior to marriage if engaged
- (c) changing religious beliefs because of what is taught in a course
- (d) breaking living unit regulations regarding hours
- (e) social drinking

Split half reliabilities were calculated for each scale. These corrected reliabilities were as follows: Daughter Wrong .87; Perceived Mother Wrong .84; Not Done, .88.

The MMPI and the Moral Ethical Values Scales were administered to a sample of 136 college females. The sequence in which the Moral Ethical Value Scales were administered was alternated in order to reduce the influence of any particular sequence. All of the subjects included in the population were single, white, Protestant, of average college age who, except for the time spent at college, were living with their mothers at the time of testing. All were volunteers from second or third year courses in psychology or education. They were not asked to put their names on the test. Instead they signed their names to a sign-up sheet and code numbers were used to match tests and subjects. The volunteers did not know the matchings could be made.

The items on the Moral Ethical Values scales were sent to the mothers of these 136 subjects. They were asked to indicate whether they felt that each item described behavior which would be morally or ethically right or wrong for a female of college age. 102 mothers responded and 95 of the responses were usable. The others were discarded for failure to follow directions accurately. The mothers were not asked to sign their names to these scales and the mothers and daughters scales were matched by using code numbers.

The daughters of the 95 mothers whose responses were used in the study were compared with the daughters of the mothers who were not included. The scores received on the neurotic triads of the MMPI and on the three Moral Ethical Value scales were used for these comparisons. Means and critical

ratios were calculated to determine if there were significant differences between the means of these two groups. No significant differences were found between these two groups on the scores they received on either the neurotic triad of the MMPI or on the Three Moral Ethical Value scales.

The subjects received the following scores on the Moral Ethical Values scales:

- (a) Daughter Wrong (DW). The number of items on this scale which the daughter considered morally or ethically wrong for a female of her age.
- (b) Perceived Mother Wrong (PMW). The number of items the daughter marked indicating that she believed her mother would consider the behavior being described as morally or ethically wrong for a college female.
- (c) Not Done (ND). The number of items on the scale which described behavior in which the subject had not engaged.
- (d) Mother Wrong (MW). The number of items on the scale which the mother indicated she would consider morally or ethically wrong for a female of college age.
- (e) Daughter Behavior Conflict (DBC). The frequency with which the daughter had engaged in behavior which she indicated was morally or ethically wrong.
- (f) Perceived Mother Behavior (PMBC). The frequency with which the daughter had engaged in behavior which she perceived her mother as considering morally or ethically wrong.
- (g) Mother Behavior Conflict (MBC). The frequency with which the daughter had engaged in behavior which the mother considered morally or ethically wrong.

The subjects were divided into groups on the basis of the scores they received on the MMPI. All of the subjects rated in the top third of the sample on the *Hs* scale were placed in one group called the *Hs* group. Subjects who fell into the top third on the *D* scale were placed in a *D* group, and subjects who rated in the top third on the *Hy* scale were placed in a *Hy* group. It was possible for subjects to be included in more than one neurotic group if they had received high ratings on more than one scale. The remaining subjects, none of whom had been included in a neurotic group, were formed into a non-neurotic group.

Table 1 shows the mean *T* scores and *T* score ranges received by the neurotic and nonneurotic groups on the MMPI.

Divisions into top 3rd and bottom 2/3rds could never be exact because subjects received scores which were at the cut-off point between the upper third and bottom two thirds. This was resolved by including all subjects who tied at this cut-off point in the nonneurotic group. The number of subjects in each group is as follows: *Hs*, 35; *D*, 39; *Hy*, 44; Non-neurotic, 67. Only 95 of the mothers of the total of 136 subjects were used in the study. Thus, the size

TABLE 1
MEAN *T* SCORES (ROUNDED) AND *T* SCORE RANGES RECEIVED BY THE NONNEUROTIC AND NEUROTIC GROUPS ON THE NEUROTIC TRIAD OF THE MMPI

Group	Mean <i>Hs</i> <i>T</i> Score	<i>Hs</i> <i>T</i> Score Range
Nonneurotic	46	28-52
Hypochondriacal	59	53-72
Nonneurotic	43	32-51
Depressive	60	52-76
Nonneurotic	50	40-57
Hysterical	65	58-80

of each of the above groups was reduced when the MW scores were used. The number in each group was reduced to: *Hs*, 24; *D*, 24; *Hy*, 31; nonneurotic, 48.

Means for these four groups were calculated for the scores received on the Moral Ethical Values scales. Each mean score received by the neurotic groups was compared with the means received on the same scale by the nonneurotic group. Critical ratios were calculated to determine whether the differences were significant.

RESULTS

Table 2 shows the mean scores received by the neurotic and nonneurotic on the Moral Ethical Values scales. Critical ratios have been calculated for the difference between

TABLE 2
MEANS AND THE CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THESE MEANS OF THE SCORES RECEIVED BY THE NONNEUROTIC AND NEUROTIC GROUPS ON THE MORAL-ETHICAL VALUES SCALES

Group	DW	ND	PMW	MW
Nonneurotic	37.04	28.60	47.81	51.48
Hypochondriacal	36.01	28.58	44.72	47.81
Critical ratio	.53	.00	1.82	1.94
Nonneurotic	37.04	28.60	47.81	51.48
Depressive	35.31	28.15	45.43	47.87
Critical ratio	.97	.21	1.23	1.82
Nonneurotic	37.04	28.60	47.81	51.48
Hysterical	34.75	28.66	44.16	47.00
Critical ratio	1.29	.00	2.50*	2.73**

* Significant at 5% level of confidence.

** Significant at 1% level of confidence.

these means in order to evaluate the significance of the differences.

Table 2 shows that there are no significant differences between the DW scores of neurotic and nonneurotic groups. A critical ratio of only 1.29, which is not large enough to be significant, was calculated for the difference between these two means. Behavior standards then did not distinguish the neurotic from the nonneurotic group. The slight differences which were found were in the direction of the neurotic groups having lower scores on the DW scale.

There were no significant differences between the ND scores of the neurotic and nonneurotic groups. Thus, there were no differences in the reported behavior of these two populations in situations involving a moral or ethical judgment.

The subjects in the *Hy* group received lower PMW scores than did the subjects in the nonneurotic groups. This difference reached the 5% level of significance. The *HS* and *D* groups also perceived their mothers as considering fewer items wrong, but these differences were not large enough to be significant; although in the case of the *HS* group, the difference did reach the 10% level of confidence.

In all three comparisons, the mothers of subjects in the neurotic groups considered fewer items wrong than did mothers of the nonneurotics. All of these differences reached the 10% level of significance, and when the mothers of the subjects in the *Hy* group were compared with mothers of the subjects in the nonneurotic groups, this difference reached the 1% level of significance.

Thus, both the mothers and the perceived mothers of the hysterical group considered a significantly fewer number of the items wrong than did the mothers of the nonneurotics. The same tendency was found when PMW and MW scores received by the other neurotic groups were compared with the nonneurotic group, but the differences were not large enough to be significant above the 10% level of confidence.

The data in Table 3 are presented in terms of proportions instead of in actual frequencies. This was done because all of the conflict scores are in part dependent upon the DW,

TABLE 3
PROPORTION OF THE ITEMS ON THE DW, PMW, AND MW SCALES MARKED AS MORALLY OR ETHICALLY WRONG BUT WHICH DESCRIBED BEHAVIOR IN WHICH THE SUBJECT HAD ENGAGED

Group	DBC	PMBC	MBC
Nonneurotic	35	46	49
Hypochondriacal	36	46	47
Critical ratio	.10	.28	.15
Nonneurotic	35	46	49
Depressive	35	48	51
Critical ratio	.00	.20	.16
Nonneurotic	35	46	49
Hysterical	35	43	45
Critical ratio	.00	.30	.35

Note.—Critical ratios have been calculated to determine if these conflict scores are significantly different for the neurotic and nonneurotic groups.

PMW, and MW scores. For example, a subject having a high MW score would have more opportunity to come in conflict with the maternal standards than a subject with a lower score. Since there were differences in the DW, PMW, and MW scores of the neurotic and nonneurotic groups, differences between the proportions of conflict rather than an actual count of the number of conflicts give more meaningful results.

Table 3 shows no significant differences between the conflict scores received by the neurotic and nonneurotic groups. This indicates that the neurotic subjects do not tend to adhere more closely to their own standards, their perceived maternal standards, or the maternal standards than do the nonneurotic subjects.

DISCUSSION

These results either do not uphold, or are directly antithetical to the impulse repression theory of anxiety which was previously presented. In no instance did a neurotic group receive a higher PW, PMW, ND, or MW score indicating greater inhibition of behavior than did the nonneurotic group. When differences that approached or reached a significant level were found, the neurotic groups received lower scores indicating less behavior restriction. Nor did the conflict scores indicate that the neurotics were more controlled

by their own standards, the perceived mother's standards, or the mother's standards than were the nonneurotics.

The basic question raised by this research is "Why were low PMW and DW scores associated with Neurosis?" Perhaps an explanation can be found in Fenichel's (1945) discussion of what he refers to as the heteronomous super ego or in Mowrer's (1950) discussion of super ego repression. Both Fenichel and Mowrer have emphasized the effects of the individual's confusion about standards of behavior. It is possible that the liberal PMW and MW scores received by the mothers and the perceived mothers were actually indications of confusion rather than of liberality, and further that confusion is related to maladjustment.

In order to evaluate this confusion hypothesis a confusion score was calculated and correlated with the *Hy* score of the MMPI. The *Hy* scale was selected because it was most closely associated with low PMW and MW scores. The confusion score was calculated by counting the number of times the daughter had misperceived the mother's standards. The Pearson Product-Moment Correlation between the frequency with which the daughter had misperceived her mother's standards and hysteria was .30. This correlation is significant at the 1% level of confidence. Thus the daughter's confusion about maternal standards was correlated with a measure of neuroticism.

This research introduces questions which need further exploration. The first step might be a replication of the experiment reported here. If at all possible, neurotic groups should be selected which are more deviant than the subjects included in this study. Also, subjects should be allowed to respond to the items with a scaled response rather than with a dichotomous response. Such scaling would allow an evaluation of the confusion hypothesis which was already presented.

SUMMARY

The objective of this research was to test the impulse repression theory of anxiety. A scale of behavior standards called the Moral Ethical Value scale and the MMPI were ad-

ministered to a population of 136 single, white, Protestant college females. On the first scale of the Moral Ethical Values scale, the subjects were asked to indicate whether they felt the behavior described by each item would be morally or ethically right or wrong for a college female. On the second scale they were asked to indicate whether they felt their mothers would consider the behavior described as morally or ethically right or wrong for a college female, and on the third scale they were asked to indicate whether they had engaged in the behavior being described. Scales were also sent to the mothers who were asked to indicate whether they felt the behavior described would be morally or ethically right or wrong for a college female.

Subjects were divided into depressive, hysterical, hypochondriacal and nonneurotic groups on the basis of scores received on the neurotic triad of the MMPI.

The neurotic and nonneurotic groups were compared on the scores they received on the Moral Ethical Values scales. These scales gave a measure of the mothers' and daughters' evaluation of behavior in situations involving moral or ethical judgment. The results showed no significant differences between the neurotic and nonneurotic group in the number of items on the scale that they indicated were morally or ethically wrong, nor was the behavior of the neurotic group different from the behavior of the nonneurotic group in situations involving a moral or ethical judgment. The neurotic daughters tended to perceive their mothers as considering fewer items wrong than did the nonneurotics. Also the mothers of the neurotics tended to consider fewer items wrong than did the mothers of the nonneurotics. Both the perceived mothers and the mothers of the *Hy* subjects considered significantly fewer items wrong than did the mothers of the nonneurotic subjects. There was no difference in the proportion of time the neurotics and nonneurotics engaged in behavior which conflicted with their own, the perceived mothers, or the mother's behavior standards.

The results of this research are antithetical to the impulse repression theory of anxiety. From this theory one would predict greater restriction of behavior to characterize the

neurotic groups. In some comparison the direct opposite of these predicted results was found.

The hypothesis was presented that liberality was actually a reflection of confusion about moral and ethical standards. In order to evaluate this hypothesis, the frequency with which the daughter misperceived her mother's standards was correlated with the hysteria scale of the MMPI. A significant correlation of .30 was calculated. This tended to support a confusion hypothesis.

REFERENCES

- BROWN, F. An experimental study of parental attitudes and their effect upon child adjustment. *Amer. J. Orthopsychiat.*, 1942, 12, 224-230.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- FREUD, S. *The problem of anxiety*. New York: Norton, 1936.
- GUILFORD, J. J. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1942.
- LASKO, JOAN. Parent-child relationships. *Amer. J. Orthopsychiat.*, 1952, 22, 300-306.
- MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.
- STONE, CAROL, & LANDIS, PAUL H. An approach to authority patterns in parent-teen age relationships. *Rur. Sociol.*, 1953, 18, 223-242.
- SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington, D. C.: William Alanson White Psychiatric Foundation, 1947.
- WATSON, G. A comparison of the effects of lax versus strict home training. *J. Soc. Psychol.*, 1934, 5, 102-105.
- WATSON, G. Some personality differences in children related to strict or permissive parental discipline. *J. Psychol.*, 1957, 44, 227-249.

(Received July 31, 1959)

CRITICAL FLICKER FREQUENCY AND THE SPIRAL AFTEREFFECT WITH PROCESS AND REACTIVE SCHIZOPHRENICS¹

JOSEPH M. McDONOUGH²

Michigan State University

Considerable theory and some research suggest that at least certain schizophrenics suffer from a disease that is primarily organic in nature, perhaps of a central nervous system locus. Traditionally it has been the tacit assumption of many workers in the area that schizophrenia has an organic etiology, most likely involving a disturbance of the brain itself (Bleuler, 1950; Noyes, 1953). The number of published findings to support such an hypothesis are far outnumbered by articles which merely speculate on such a possibility without any attempt at experimentation or controlled observation. A few studies nevertheless do offer some support for the assumption that brain damage is the important factor in schizophrenia. Evidence from biopsies obtained from 70 cases of dementia praecox led Bateman and Papez (1951) to suggest that changes in nerve cells of the prefrontal cortex are due to pathogenic organisms. From an examination of 1087 records, Kallmann (1938) concluded that "the organic process of schizophrenia in the brain . . . attacks the cortex almost exclusively and usually leaves the internal parts of the brain intact." Kennard and Levy (1952) found abnormal EEGs, often indicative of brain lesions, significantly more often in the more severe

cases of schizophrenia, especially those resembling dementia praecox. Winkelman and Book (1948) concluded on the basis of 10 autopsies of schizophrenics that gross brain appearance did not differ from normals but that microscopic evidence did suggest the diagnosis.

Widespread dissatisfaction with the traditional Kraepelinian categorization of schizophrenia or dementia praecox has led to increased effort in recent years to refine the concept. One suggestion has been to view the disorder as including two polar types. At one extreme is the patient who has an early and insidious onset of psychosis with absence of precipitating stress, an inadequate prepsychotic personality with a tendency to withdraw from all human contacts, clinical symptoms showing a flat or dull affect, and relative absence of confusion. This syndrome has been called *process* schizophrenia. *Reactive* schizophrenia on the other hand is the term given to describe the case where there is typically an abrupt and stormy onset of the psychosis usually attributable to a logical and significant stress situation. Here the premorbid personality may be described as normal or neurotic, rather than schizoid, with perhaps some degree of outgoingness present and a clinical picture likely to include severe confusion and many affective components. Attempts to classify schizophrenic samples into the two categories or to view the individual patient as occupying a position on a process-reactive continuum have increased in number and a certain degree of success has been achieved in differentiating the groups experimentally (Becker, 1956; Kantor, Wallner, & Winder, 1953; King, 1958). One investigation by Brackbill and Fine (1956) sought to

¹ This research is based on a doctoral dissertation done at Michigan State University. The author wishes to express his appreciation to M. Ray Denny, Gerald F. King, S. Howard Bartley, Alfred Dietze, and G. M. Gilbert for their many contributions to this thesis. Particular thanks are due also to Stewart G. Armitage and his staff at the VA Hospital, Battle Creek, Michigan for their continuous cooperation in making this study possible.

² Now at the Veterans Administration Hospital, Coral Gables, Florida.

relate the concepts specifically to central nervous system pathology. Although it was impossible to differentiate the process from the organic group on the basis of Rorschach indices of brain damage, it was observed that both of these groups had significantly more individuals with five or more of the "organic signs" than did the reactive group. The authors suggested, therefore, that "the difficulty in differential diagnosis of some kinds of schizophrenics and organics results from the involvement of central nervous system pathology in process schizophrenia."

Two perceptual techniques of "brass instrument" psychology have been found fairly effective in differentiating brain damaged from nonorganic groups: critical flicker frequency (CFF) and the Archimedes Spiral. Numerous studies have substantiated the fact that brain damaged patients have significantly lower CFF thresholds than do nonorganic patients (Ross & Reitan, 1955; Teuber & Bender, 1948; Werner & Thuma, 1942). In psychiatric patients with functional disorders CFF has not been investigated as thoroughly, and most studies are open to methodological criticisms including the small number of patients involved. In two of the better controlled studies, however (Irvine, 1954; Ricciuti, 1948), the preponderance of evidence indicated that schizophrenic patients have significantly higher CFF thresholds than do brain damaged patients and that they do not differ significantly from normal individuals.

Visual fixation by *S* upon a slowly rotating Archimedes Spiral produces the illusion of expansion or contraction of the disc depending on the direction of rotation. When the spiral is brought to a sudden stop the perceiver experiences the illusion just the reverse of that during rotation, i.e., a negative aftereffect. Although the extremely high discrimination achieved by Price and Deabler (1955) in differentiating brain damaged from nonorganic patients has not been replicated, subsequent studies generally have found that significant differences do exist between the two groups and that the technique is effective to some extent as a clinical test of organicity (Gallese, 1956; Page, Rakita, Kaplan, & Smith, 1957). On the assumption, therefore, that process schizophrenia involves central nervous sys-

tem pathology and that this is specifically cortical in nature, it was hypothesized that process schizophrenics and brain damaged patients would have significantly lower CFF thresholds and would fail to perceive the spiral aftereffect significantly more often than would reactive schizophrenics and normals.

METHOD

Apparatus

The CFF was tested by means of an episcotister viewed monocularly by *S*. Visual angle of the target was 2 degrees 28 minutes with an intensity of .085 candles per square foot. A shutter which could be lowered over the target aperture allowed *E* to terminate the stimulus immediately at the end of each flicker trial. Power for the episcotister was supplied by a Cenco 115-volt, 60-cycle motor. Rate of rotation for the disc was controlled manually by *E*'s turning a small knob on the motor. Four equal sized sectors were cut in the disc in a manner to produce a light-dark ratio of 1.0. An automobile speedometer was used to measure the speed of the spinning disc and calibration by a Strobotac proved linear.

The Archimedes Spiral was made from a cardboard disc, off-white in color and 8 inches in diameter. A black stripe, one-half inch wide, was painted on the cardboard so that it made one and three quarter turns from the center until it tapered completely off the outside edge of the disc. Speed of rotation was kept constant at 100 rpm.³ When rotated, an illusion of the disc contracting or drawing into the center was created in the typical viewer. On stopping the disc suddenly the "normal" experience was to perceive the spiral as expanding.

Subjects

Eighty hospitalized male veterans between 19 and 45 years of age with good visual acuity were placed in four groups of 20 *S*s each. The organic group consisted of individuals with *known* brain damage, chosen on the basis of their diagnosis as determined by neurological findings of the medical staff, case history material, and presenting symptomatology. Diagnoses included multiple sclerosis, brain trauma, brain tumor, Huntington's Chorea, Parkinson's Disease, CVA, and intracranial infection. Mean age was 35.2 years with patients averaging 9.3 years of schooling. The normal group was made up of 20 veterans from a general medical and surgical hospital and included every man on a surgical ward under the age of 45 with no history of emotional difficulties necessitating hospitalization or outpatient treatment. Diagnoses included postoperative fractures, appendectomies, torn cartilage, hemorrhoids, and sacroiliac

³ Some unpublished evidence appears to indicate that size of disc and speed of rotation affect *S*'s perception of the afterimage.

strain. This group averaged 31.1 years of age with a mean educational level of 10.7 years.

The third and fourth groups were made up of diagnosed schizophrenics from an NP hospital. One hundred and sixty-one schizophrenic case histories were examined. From these a total of 76 were selected on the basis of their having: a case history with sufficient material to allow rating; no evidence either behaviorally or from the record of any head injury or long period of unconsciousness; a background free from repeated NP hospitalization during the previous 10 years; and definite evidence from the history of gross schizophrenic symptomatology, e.g., hallucinations, delusions, ideas of reference, and bizarre behavior. These 76 patients were then rated on Becker's (1955) revision of the Elgin Prognostic Scale by two clinical psychologists. Interrater reliability was indicated by a Pearson r of .89. The two numerical ratings on each patient were then averaged and the resulting 76 scores ranked on a continuum. The 20 patients receiving the lowest point totals were selected as being most "reactive," and those 20 with the highest scores most "process." The decision to use Becker's rating scale was based on the belief that it offered a greater degree of exactness in the ratings, permitted a grading along the entire length of the continuum rather than a forced choice dichotomy or the creation of a middle or mixed group, and appeared to fit more easily into a theoretical framework such as Bellak's (1949) where schizophrenics are seen as ranging from almost completely organic to almost completely psychogenic.

The *process* group had a mean age of 26.6 and an educational level of 10.7 years, while the *reactive* group averaged 33.6 years of age and had 10.8 years of schooling. Mann-Whitney U tests revealed no significant differences among the four groups in terms of educational level, but significant differences in age were obtained. Process schizophrenics were significantly younger than the other three groups ($p < .001$), a characteristic consistent with the early insidious onset associated with this group. Organics were significantly older than normals ($p < .05$), which may be partly attributable to the progressive nature of some of the organic disorders, allowing the patients to remain unhospitalized until somewhat later in life.

Procedure

All Ss were informed that they were to undergo two "special eye tests." With the Archimedes Spiral, S were seated 8 ft. from the face of the disc with the room under ordinary illumination. Instructions used were those of Gallese (1956) in which the patient was questioned as to his perception of the rotating disc. Four 30-sec. trials were given at the end of which the disc was stopped abruptly. S was then asked to describe what he "saw." A scoring system devised by Price and Deabler (1955) was used to record S's response. If his answers suggested that he perceived the aftereffect, he was given a score of 1 for that trial. If he did not report the illusion of

expansion, a score of 0 was recorded. S's final score was based on all four trials. A point total of 4, therefore, indicated that S perceived the aftereffect on all four trials, whereas a score of 0 meant that he failed to perceive the effect on any of his trials.

In the case of CFF, S was carefully instructed on how to view the patch and given the following directions:

The spot you see there [i.e., the test patch] is going to flicker off and on. Stare at it steadily and don't blink. After a while the light will no longer appear to be flickering. Instead, as you stare at it, it will gradually appear to become a steady light. As soon as it looks as though it is no longer flickering, you say, "Now," . . . and I will turn it off.

The overhead light was then extinguished and the only illumination came from the well-shielded 7½-watt bulb in the rear of the apparatus which permitted dial readings. After one minute of dark adaptation, the shutter was raised allowing S to view the test patch. E manually increased the disc speed at a constant rate (approx. 1 mph per second) until S reported the appearance of a steady light. A total of 10 successive ascending trials were taken with about 30 seconds between trials. Possible auditory cues from the episcotister were masked by an electric fan whose blades struck a small piece of cardboard attached to the fan guard. After considering the first three trials as practice, mean flicker threshold was determined for each of the 80 Ss. Patients were found to be quite reliable in their reporting of fusion with only 12 Ss obtaining an intertrial SD exceeding three quarters of a cps and only one S in the entire sample with an SD more than 1 cps.

RESULTS

Critical Flicker Frequency

Mean CFF scores for each of the four groups are summarized in Table 1. Since variances were not homogeneous, the Kruskal-Wallis H Test was used. An H value significant at the .001 level of confidence was obtained indicating that the four groups differed in mean flicker threshold beyond chance expectation. Mann-Whitney U Tests run be-

TABLE 1
CRITICAL FLICKER FREQUENCIES
IN CYCLES PER SECOND

	Mean	Variance
Organic	13.20	2.30
Normal	16.38	.26
Process	16.43	.73
Reactive	16.67	.94

tween all combinations of the groups in pairs showed that the organic group differed very significantly ($p < .0001$) from each of the other three groups. A significant difference in CFF at the .05 level was found to exist also between normals and reactives. Although mean CFF for the process group was slightly lower than that of the reactive group, the difference was not significant. In order to determine whether a more rigorous dichotomization would yield differences between the two schizophrenic groups, the 10 Ss rated most reactive (i.e., lowest Elgin ratings) were compared with the 10 rated most process (i.e., highest ratings). Again, no significant differences were found between process and reactive patients in terms of CFF. On the possibility that a relationship existed between a schizophrenic S's flicker rank and his rank on the Elgin Scale, Kendall's tau was computed. The resulting correlation of .048 indicated essentially no relationship between the magnitude of the schizophrenic's Elgin rank (or position on the process-reactive continuum) and his ranked CFF score. The ability of CFF to discriminate brain damaged from all other patients tested was exceptionally good, however. When compared with normals, organic Ss were successfully discriminated by their flicker threshold 95% of the time since only one organic S obtained a higher CFF score than that of the lowest normal.⁴ Somewhat less discrimination was present in comparison of process and reactive groups with the organics (90% and 75% respectively).

Archimedes Spiral

With the exception of the 20 brain damaged patients, all of the Ss (i.e., 20 normals, 20 process, and 20 reactives) obtained perfect scores on the task. Although χ^2 revealed that scores on the task. Although χ^2 revealed that organics were significantly different from each of the other groups ($p < .01$) the ability of the spiral technique to discriminate brain injured subjects from nonorganic individuals was relatively poor. Of the 20 organics only 9 failed to obtain perfect scores on the task. This resulted in a test discrimination index of 45%. Only 5 of the 9 organics were totally unable to perceive the effect on any of the four trials. No relationship was observed

between organics' scores and their neurological diagnoses.

DISCUSSION

There were no differences in CFF between schizophrenics classified process and reactive. Although the method of classification may be speculated on as one reason for this failure, other studies utilizing the method have differentiated the two groups (Becker, 1955; King, 1958). Present results do not suggest demonstrable cortical defect as measured by CFF in either process or reactive schizophrenia. The fact that the reactive group had a higher CFF than did the normals warrants speculation. Heightened physiological responsiveness of reactive schizophrenics noted in previous studies was attributed by King (1958) to the heightened homeostatic response which reflects the precipitous nature of their adjustment. Other investigations (Ricciuti, 1948; Wiersma, 1906) have indicated that manic patients have higher CFFs than do other groups. Similarity between manic or affective symptoms and characteristics of schizophrenics who might be classified reactive has also been noted in previous studies (Cameron, 1938; Funkenstein, Greenblatt, & Solomon, 1948). Affective components in the clinical picture are among the bases for classification as reactive in the rating scale employed in the present investigation. Therefore, it seems possible that some patients who formerly might have been diagnosed manic-depressive are now being diagnosed schizo-affective or reactive schizophrenic, and these manic features may account for the reactive group's higher than normal CFFs. On the other hand, the significant difference found between normals and reactives may merely indicate a comparison of the extremes in a homogeneous distribution.

A possible alternative explanation of the results might contend that a defect in the experimental design was responsible for lower flicker thresholds in the organics. This position would maintain that since only ascending series were used in the present study it is possible that the organics were responding prior to their subjective impressions of complete fusion and that this premature reaction would also have been obtained in descending

⁴ This S was a frontal lobe injury.

series, thus changing the results. However, since previous investigations (Irvine, 1954; Teuber & Bender, 1949; Werner & Thuma, 1942) utilizing both ascending and descending series have yielded consistently lower thresholds for organics, it would seem more parsimonious to accept a lower CFF threshold on the part of organics, rather than premature reaction, as the explanation for these results.

The present experiment appears to provide additional support for CFF as a good technique in discriminating brain damaged persons with gross destruction from normals and schizophrenics. Increased utilization of the technique in the ordinary hospital setting seems to be well warranted, a possibility recognized by other investigators. Its potential seems especially good when a comparison is made with the relatively poor discriminating ability shown by the usual paper and pencil tests of organicity.

Although the spiral aftereffect differentiated organic Ss from all other groups, only 45% of the brain damaged Ss were identifiable through failure to perceive the effect. Since only known organic cases were included in the present study it seems unlikely that this technique will be of extensive diagnostic value in a clinical setting where less severe organic cases are to be classified. The fact that none of the schizophrenics of either type failed to perceive the aftereffect does not, at least, support the position that cortical defect is involved in process schizophrenia.

SUMMARY

The present experiment was designed to investigate the possibility of organic involvement, particularly of a cortical nature, in process schizophrenia. Recent studies dichotomizing schizophrenics into process and reactive types have suggested cortical damage in the former group. It was hypothesized, therefore, that patients classified as process schizophrenics would perform similarly to brain damaged individuals on two perceptual tasks, while reactive schizophrenics would more closely resemble normal controls. The two perceptual tasks used were critical flicker frequency and the Archimedes Spiral aftereffect,

both of which had proven effective previously in differentiating brain damaged from non-organic Ss. Four groups of 20 Ss were given both tasks: reactive and process schizophrenics, organics with known brain damage, and normals. Results of the experiment showed that organic patients were significantly different from all other groups in terms of CFF threshold and ability to perceive the spiral aftereffect. Process and reactive schizophrenics did not differ from each other on either task. Reactive schizophrenics differed from normals on CFF in having higher thresholds. Attempts to account for the findings included the possibility that affective components of reactive schizophrenia contribute to the difference between them and other schizophrenics and normals, and the belief that process schizophrenia, while showing some physiological differences from reactive schizophrenia in other investigations, probably is not related to defects of a cortical nature.

REFERENCES

- BATEMAN, J., & PAPEZ, J. Significance of the thalamus in psychoses with reference to hallucinations and delusions. *J. clin. exp. Psychopath.*, 1951, 12, 89-103.
- BECKER, W. The relation of severity of thinking disorder to the process-reactive concept of schizophrenia. Unpublished doctoral dissertation, Stanford Univ., 1955.
- BECKER, W. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, 53, 229-236.
- BELLAK, L. A multiple factor psychosomatic theory of schizophrenia. *Psychiat. Quart.*, 1949, 23, 738-755.
- BLEULER, E. Dementia praecox or the group of schizophrenias. New York: International Univ. Press, 1950.
- BRACKBILL, G., & FINE, H. Schizophrenia and central nervous system pathology. *J. abnorm. soc. Psychol.*, 1956, 3, 310-313.
- CAMERON, D. Early schizophrenia. *Amer. J. Psychiat.*, 1938, 95, 567-582.
- FUNKENSTEIN, D., GREENBLATT, M., & SOLOMON, H. Autonomic nervous system changes following electric shock treatment. *J. nerv. ment. Dis.*, 1948, 108, 409-422.
- GALLESE, A. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, 12, 254-259.
- IRVINE, R. CFF for paretics and schizophrenics. *J. abnorm. soc. Psychol.*, 1954, 49, 87-88.
- KALLMANN, F. The genetics of schizophrenia. New York: Augustin, 1938.

- KANTOR, R., WALLNER, J., & WINDER, C. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, 17, 157-162.
- KENNARD, MARGARET, & LEVY, S. The meaning of abnormal electroencephalograms in schizophrenia. *J. nerv. ment. Dis.*, 1952, 116, 413-423.
- KING, G. Differential autonomic responsiveness in the process-reactive classification of schizophrenia. *J. abnorm. soc. Psychol.*, 1958, 56, 160-164.
- NOYES, A. Modern clinical psychiatry. Philadelphia: Saunders, 1953.
- PAGE, H., RAKITA, G., KAPLAN, H., & SMITH, N. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, 21, 89-91.
- PRICE, A., & DEABLER, H. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, 19, 229-302.
- RICCIUTI, H. A comparison of critical flicker frequency in psychotics, psychoneurotics, and normals. *Amer. Psychologist*, 1948, 3, 276.
- ROSS, A., & REITAN, R. Intellectual and affective functions in multiple sclerosis: A quantitative study. *AMA Arch. Neurol. Psychiat.*, 1955, 73, 663-677.
- TEUBER, H., & BENDER, M. Changes in visual perception of flicker, apparent motion, and real motion after cerebral lesion. *Amer. Psychologist*, 1948, 3, 246-247.
- TEUBER, H., & BENDER, M. Alterations in pattern vision following trauma of occipital lobes in man. *J. gen. Psychol.*, 1949, 40, 37-57.
- WERNER, H., & THUMA, B. Critical flicker frequency in children with brain injury. *Amer. J. Psychol.*, 1942, 55, 394-399.
- WIERSMA, E. Die Sekundärfunktion bei Psychosen. *J. Psychol. Neurol. Lpz.*, 1906, 8, 1-24.
- WINKELMAN, N., & BOOK, M. Observations on the histopathology of schizophrenia: I. The cortex. *Amer. J. Psychiat.*, 1948, 105, 889-896.

(Received April 16, 1959)

AN ATTEMPT TO EMPLOY A SENTENCE COMPLETION TEST FOR THE DETECTION OF PSYCHIATRIC PATIENTS WITH SUICIDAL IDEAS¹

HERMAN Y. EFRON

Veterans Administration Hospital, Lyons, New Jersey

Construction of a test suitable for group administration which could detect those who are morbidly preoccupied with depressive thoughts would be extremely useful to ward administrators in hospitals where intensive psychotherapeutic contacts are the exception rather than the rule. This study is an attempt to assess a group screening device (sentence completion test) to identify those patients who may constitute suicidal risks.

The *D* scale of the MMPI (Broida, 1954) and some aspects of responses to the Make-a-Picture-Story (MAPS) (Farberow, 1950) differentiated between suicidal and nonsuicidal patients. However, the few experimental studies with the TAT yielded negative results. Neither responses to Card 3BM (Broida, 1954) nor characteristics of the TAT heroes (Shneidman & Farberow, 1958) distinguished suicidal from control groups. There have been proposed in the literature various Rorschach signs and patterns that can be used to detect potential suicidal behavior (Hertz, 1948, 1949; Piotrowski, 1950; White & Schreiber, 1952); however, other studies generally fail to confirm their validity (Fisher, 1951; Sakheim, 1955). Results indicate that Hertz's (1948) Configuration Technique is probably

the most useful of the prognostic scales (Sakheim, 1955), but even her technique failed when used with schizophrenic patients (Fisher, 1951). Even if such a test as the Rorschach were capable of differentiating between potential suicidal patients and nonsuicidal patients, the time involved in arriving at these decisions would preclude their general use. The number of psychologists needed to service a psychiatric hospital would be far in excess of current staffing patterns even though periodic retesting was restricted to those patients whose case histories warrant some suspicion of self-destructive behavior. A self-administering test which could be repeated at specified intervals, tapping both attitudes within and outside the patient's awareness, seemed to be the most desirable instrument. It was felt that the sentence completion test might be such an instrument.

The most difficult problem in designing an experiment to test the efficiency of an instrument to predict suicide is that of an appropriate criterion. Because of the low incidence of suicide it is not practicable to rely upon this as a criterion measure. Rosen (1954) has clearly analyzed the inherent problems in predicting an event whose occurrence among psychiatric patients (as judged from statistics covering 1936 to 1949 at a large state hospital) is roughly 3 per 1,000 admissions² (Levy & Southcombe, 1953). In his article, Rosen (1954) demonstrates that if a test were

¹ This study was an individual hospital project of the Veterans Administration Psychiatric Evaluation Project, Richard L. Jenkins, Director. This paper was reviewed in the Veterans Administration and is published with the approval of the Chief Medical Director. The statements and conclusions of the author are the result of his own study and do not necessarily reflect the opinion or policy of the Veterans Administration. The author wishes to express his gratitude to Marcus P. Rosenblum for his constructive suggestions, and to Leon Cohen, Eric Layne, David Levine, and John Seymour for their cooperation in rating the sentence completions.

² Using the same data, Rosen (1954) incorrectly arrived at .0032 as a rough index of suicide among the psychiatric population. Actually, since there are about three to four times as many resident patients per year as admissions per year, a closer approximation to the actual suicide rate would be one third to one fourth of .0032.

devised that could correctly identify whether a patient was suicidal or nonsuicidal 75% of the time, the number of false positives (i.e., nonsuicidal patients being identified as suicidal) in a patient population of 12,000 would be close to 3,000. Recognizing this severe limitation it was still felt that if a group screening device that would approach 75% accuracy could be developed it would serve to alert the hospital personnel to investigate more thoroughly all patients designated as "suicidal." In a hospital of 2,000 this would mean more vigilant study of about 500 patients, an aim that might be feasible. Rosen (1954) emphasizes the need for researchers to formulate more precise definitions of "suicidal" patients. Studies by Farberow (1950) and Rosen, Hales, and Simon (1954) found marked group differences between the MMPI scores of patients who attempted suicide and those who threatened suicide. The group that made threats gave significantly higher mean scores on most all scales. Both studies indicate that as a group those who made suicidal threats were more severely disturbed (at the time of testing, which was after those in the attempted suicide group had executed their attempts). The present study deals with patients who were known to have expressed conspicuous suicidal thoughts, marked nihilistic ideas, or severe guilt with concern over punishment or torture. A basic assumption is that patients with such morbid ruminations are greater suicidal risks than those who do not have such preoccupations. The only evidence that can be cited to support the assumption bears indirectly upon it and is possibly irrelevant, viz., Shneidman and Farberow's (1956) report that almost all of the people they studied who did commit suicide had given some earlier warning of their intentions. The hypothesis of this study is that psychologists can, on the basis of clinical judgment, differentiate between the sentence completions of patients with suicidal ruminations and those without such ruminations.

METHOD

Test Instrument

Some incomplete sentences were purloined from existing tests and still others were constructed until

the finished test consisting of 54 items was evolved.³ Almost a third of the items appear to be weighted toward eliciting depressive themes. Included in this group were such items as "If there were no way out I would—," "My greatest fault is—," and "Bob's defeat made him—."

Procedure

Subjects (Ss) were tested in small groups of 10 to 18 patients were given one hour in which to complete the test. Throughout the testing session the examiner encouraged those patients who seemed to be having difficulty in completing the sentences. When more than about half the items were omitted the patient was asked to write the reason for his difficulty in completing the sentences. In a few instances the patients even refused to do this, and the examiner wrote the reason the patient stated verbally.

Phase I. Of the 104 obtained completions 54 were given to four staff psychologists. They were told that 18 were done by suicidal patients, 9 by assaultive patients, 27 by patients who were neither suicidal nor assaultive, and they were then asked to distribute them into their proper categories. After every judgment they were requested to state the degree of confidence they had in their judgment. Coding for these statements was: 1. I'm in doubt whether this is correct. 2. I feel fairly confident that this correct. 3. I feel confident that this is correct. The only identifying data were the patient's code number and age.

Phase II. From the 54 completions used in Phase I, all those with less than 25 responses were eliminated. The same four psychologists were asked to redistribute the tests which now represented 10 suicidal patients, 3 assaultive patients, and 25 nonsuicidal and nonassaultive patients. They were again asked to rate their confidence in each decision.

Phase III. In order to obtain an estimate of the frequency of false positives, the remaining 50 completions (which were all done by patients who were nonsuicidal and nonassaultive) were given to three of the four psychologists with the following instructions:

The following sentence completion tests were obtained from patients who may have been on suicidal observation status, on assaultive status, on *suicidal and assaultive* status, or on none of the above. On the accompanying Summary Sheet, will you please distribute the code numbers of the patients into their appropriate categories. Remember the frequency in any of the above categories may be zero.

This time there was also a "cannot rate" category in the event they found it impossible to reach a decision. Again they were requested to state the degree of confidence they had in each judgment.

Subjects. Patients were selected from the acute intensive treatment service of a large Veterans Ad-

³ A copy of the incomplete sentences can be obtained from the author.

TABLE 1
DESCRIPTION OF SAMPLE

Demographic Variable	Phase I		Phase II
	Suicidal	Nonsuicidal	Nonsuicidal
<i>N</i>	18	36	50
Diagnosis			
Schizophrenic	18	33	44
Manic-depressive	0	2	3
Chronic brain syndrome	0	1	1
Other	0	0	2
Current hospitalization			
Less than 6 months	7	15	16
Six months to 2 years	2	13	16
More than 2 years	9	8	18
All NP hospitalizations			
Less than 6 months	4	1	2
Six months to 2 years	2	12	18
More than 2 years	12	23	30
Age			
Mean	35	37	36
<i>SD</i>	5	11	9

ministration hospital. For a patient to have been considered suicidal he must have expressed serious self-destructive thoughts, marked nihilistic ideas, or severe guilt with concern over punishment and torture to the ward physician, nurse, or aide within the week prior to testing. To have been considered assaultive⁴ the patient, in the estimation of the ward physician, had to present a high risk of acting out aggressively. Those patients who, for at least two months prior to testing, had not expressed any self-destructive or depressive ideas and had no known history of suicidal attempt were considered nonsuicidal. Nonassaultive patients were those whom the ward physician had no reason to suspect would act out aggressively. Of the 104 Ss, 18 were suicidal, 9 were assaultive, and 77 were both nonsuicidal and nonassaultive. Of the latter group, 33 were privileged (i.e., had freedom of the hospital grounds) and 44 were nonprivileged. Table 1 contains a further description of the groups.

RESULTS AND DISCUSSION

The number of correctly assigned suicidal patients was compared with the chance expectancy of such an occurrence as computed

⁴ Assaultive patients were included in the study in order to approximate the clinical situation and thereby encourage each rater to evaluate the completions more critically. Also, it was intended to prevent spurious positive findings that may have resulted by virtue of the fact that more severely disturbed patients have a greater likelihood of being rated suicidal.

from the hypergeometric function. For Phase I only one of the four psychologists' distributions warranted assuming that other than chance factors were operating. The probabilities obtained were .82, .38, .06, and .02. When the tests which contained less than 25 responses were discounted (Phase II) none departed significantly from chance expectation ($P = .97, .53, .53, .06$). The over-all percentage of correct identification of patients with suicidal ruminations was 43% for Phase I and 30% for Phase II. Since in Phase III the true distribution was not disclosed to the psychologists, their ratings provide us with an index of the false positives we might expect. The mean percentage of false positives was 33%.⁵ Clearly these results indicate that the test as used is woefully inadequate to provide practical value in the detection of suicidal patients.

A comparison was made between the num-

⁵ It is possible that the set established in Phases I and II by virtue of the raters being told the number of suicidal patients to expect resulted in increasing the probability of false positives. However, for one rater Phase III preceded the others. His percentage of false positives was 30%.

ber of agreements among psychologists with the number to be expected if chance were the only factor. The expectancy derived from probability theory was compared with the obtained results by means of chi square. For the purpose of these comparisons all the distributions were dichotomized into the categories "suicidal" and "others." The chi squares for Phase I and Phase II were significant beyond the .01 level of significance, and for Phase III beyond the .05 level. When the assaultive patients were grouped with the suicidal patients the chi squares for Phase I and II were significant beyond the .001 level and for Phase III beyond the .01 level. This latter finding coupled with the previous observation that the psychologists did somewhat better when records with less than 25 responses were included tends to indicate that their ratings were more in terms of severity of illness than focus upon specific traits that may be associated with suicide. Further evidence to support this formulation is available from the comparison of the number of privileged and nonprivileged patients rated as suicidal. Although both these groups were actually non-assaultive and nonsuicidal, the likelihood of a nonprivileged patient being rated as suicidal was 60% greater. Since suicidal ruminations may be the result of varied and wholly different intrapsychic processes, our findings are not surprising.

In analyzing the relationship between the degree of confidence and correctness of decision it was found that with the exception of the one psychologist who did better than chance in Phase I there was no marked increase in accuracy when a decision was made with confidence. The exceptional rater was correct in 11 of his 14 high confidence decisions and only 4 of his 14 low confidence decisions. Such a difference would occur on a chance basis less than one time in a hundred (chi square with Yates correction). Using only the high confidence ratings of this last rater he was able to place correctly 79% of the suicidal patients. It would therefore seem that such a test as the one described may, in the hands of a proficient rater, have some usefulness in detecting suicides. Further research focused upon identifying proficient raters seems warranted.

SUMMARY

One hundred and four psychiatric patients were given a sentence completion test constructed to elicit suicidal ideation. Of the 104 subjects, 18 were known to have been preoccupied with suicidal ruminations. In one part of the experiment, four clinical psychologists categorized sentence completions as coming from suicidal, assaultive, and nonassaultive patients after having been given the frequencies in each category. In another part of the experiment, three psychologists were given different sentence completions and asked to categorize them without being told the actual distribution.

The over-all percentage of correct identification of patients with suicidal ruminations varied from 30% to 43%. The mean percentage of false positives was 33%. Although the psychologists' predictions showed no greater relationship to the criterion measure than chance expectancy, they agreed with each other more often than would be expected if chance were the only factor. This, coupled with other evidence, was taken to indicate that the psychologists' ratings were primarily in terms of severity of illness rather than focusing on specific suicidal traits. As a group, the raters' performance was not significantly better when they had high confidence in their decisions. However, the high confidence ratings of one psychologist correctly placed 79% of the suicidal patients. This was taken to indicate that there was a possibility of identifying proficient raters who could use this test successfully.

REFERENCES

- BROIDA, D. C. An investigation of certain psychodiagnostic indications of suicidal tendencies and depression in mental hospital patients. *Psychiat. Quart.*, 1954, 28, 453-464.
- FARBEROW, N. L. Personality patterns of suicidal mental hospital patients. *Genet. Psychol. Monogr.*, 1950, 42, 3-79.
- FISHER, S. The value of the Rorschach for detecting suicidal trends. *J. proj. Tech.*, 1951, 15, 250-254.
- HERTZ, MARGUERITE R. Suicidal configurations in Rorschach records. *Rorschach Res. Exch.*, 1948, 12, 3-58.
- HERTZ, MARGUERITE R. Further study of "suicidal" configurations in Rorschach records. *Rorschach Res. Exch.*, 1949, 13, 44-73.

- LEVY, S., & SOUTHCORBE, R. H. Suicide in a state hospital for the mentally ill. *J. nerv. ment. Dis.*, 1953, 117, 504-514.
- PIOTROWSKI, Z. *A Rorschach Compendium*. Utica: State Hospital Press, 1950.
- ROSEN, A. Detection of suicidal patients: An example of some limitations in the prediction of infrequent events. *J. consult. Psychol.*, 1954, 18, 397-403.
- ROSEN, A., HALES, W. M., & SIMON, W. Classification of "suicidal" patients. *J. consult. Psychol.*, 1954, 18, 359-362.
- SAKHEIM, G. A. Suicidal responses on the Rorschach test: A validation study. *J. nerv. ment. Dis.*, 1955, 122, 332-344.
- SHNEIDMAN, E. S., & FARBEROW, N. L. Clues to suicide. *U. S. Public Hlth. Rep.*, 1956, 71, 109-114.
- SHNEIDMAN, E. S., & FARBEROW, N. L. TAT heroes of suicidal and non-suicidal subjects. *J. proj. Tech.*, 1958, 22, 211-228.
- WHITE, M. A., & SCHREIBER, H. Diagnosing suicidal risks on the Rorschach. *Psychiat. Quart. Suppl.*, 1952, 26, 161-189.

(Received April 17, 1959)

FACTOR ANALYSES OF "ANXIETY" AND "NEUROTICISM" INVENTORIES

A. W. BENDIG

University of Pittsburgh

In recent years a number of different questionnaires purporting to measure "anxiety" (Cattell, 1957b; Taylor, 1953) or "neuroticism" (Cattell, 1955; Eysenck, 1956a; Winne, 1951) have been independently developed by different methodologies, and subsequent studies have shown relatively high correlations among these scales (Bendig, 1957; Holtzman, Calvin, & Bitterman, 1952; Windle, 1955). These correlations suggest that the Eysenck, Winne, and Taylor scales may all be measuring the same objective personality dimension. However, several attacks have been launched against the unidimensionality of these scales. Eysenck (1956b, 1957) argues that such questionnaires are confounding the independent dimensions of neuroticism and extraversion and his hypothesis has received some experimental verification (Bendig, 1957, 1958; Frey & Becker, 1958). A second objection to such scales has been that of Edwards (1957) who suggests that much of the variance of such scales can be attributed to social desirability or the tendency of Ss to give socially desirable responses to personality items regardless of their content and the large correlations between Edwards' Social Desirability Scale (SDS) and Taylor's MAS (Edwards, 1957, p. 33) and Cattell's Anxiety scale (Bendig, 1959b) have supported this hypothesis. However, 22 of the 39 items on Edwards' scale overlap with Taylor's MAS; and when only the 17 non-MAS items are scored on the Social Desirability Scale, Edwards found the MAS correlation substantially reduced (Edwards, 1957, p. 88), and Bendig (1959b) similarly reported that the Cattell Anxiety score correlation decreased. These findings suggest that the Edwards SDS is also not unidimensional, but measures a combina-

tion of social desirability, anxiety, and/or neuroticism factors. Parenthetically it might be noted that the SDS shows a low positive correlation with the MMPI Lie scale (Edwards, 1957, p. 44), suggesting that the SDS scores may also involve a "lie" component.

At a theoretical level, Spence (1958) argues for the construct validity of Taylor's MAS as an operational definition of "manifest anxiety," now renamed "emotionality" or "emotional reactivity," while Eysenck (1956b; 1957), as noted above, regards anxiety as a combination of introversion and neuroticism. Although Spence may be regarded as a monist in this area of personality theorizing, and Eysenck as a dualist, Cattell (1957a) is a trinitarian since he regards anxiety, neuroticism, and extraversion as three relatively independent personality dimensions and has published inventories for the measurement of both neuroticism and anxiety (Cattell, 1955, 1957b). So we are left with one, two, or three personality factors with varying identifications depending on the theoretical orientation of the scientist.

The present research was designed to investigate the number of orthogonal factors present in a battery of independent questionnaires purportedly measuring the above noted personality dimensions and to clarify the factorial composition of each of these commonly used inventories. It was assumed that the intercorrelations among the anxiety and neuroticism scores might be influenced by other contaminating factors, such as extraversion, social desirability, falsification, and omnipresent sex differences, and the inclusion of marker variables for these hypothetical factors would help to purify the factors obtained. The central question of the study was

whether, when the influence of other factors was isolated by these marker variables, a single general factor of "emotionality" (Spence, 1958) or separate group factors of "anxiety" and "neuroticism" (Cattell, 1957b) would appear necessary to account for the interrelationships among the scales.

STUDY I

Procedure

Scales. The scales included in this study were assembled into a single booklet containing 230 items from which 12 scores were obtained. Only 10 scores were used with each of two groups of Ss, since four of the scores were not experimentally independent (see below).

1. Manifest Anxiety Scale (Taylor, 1953) included only the 50 dichotomous-response items that are actually scored for this scale (McCreary & Bendig, 1954).¹

2. Winne's (1951) Neuroticism scale: Since 8 of the 30 dichotomous-response MMPI items on this scale are included in the Manifest Anxiety scale, only the 22 experimentally independent items were used.²

3. Edwards' (1957) Social Desirability Scale: Only the 17 dichotomous-response MMPI items that are not included in Taylor's MAS were scored for this scale.³

4. Eysenck's (1956a) Extraversion scale, which is included in the Manifest Anxiety Inventory, contains 24 items.

5. Eysenck's (1956a) Neuroticism scale is the second 24-item scale included in the MPI. Dichotomous item responses (Yes-No) were used with items on both of the Eysenck scales (Bendig, 1959a).

6. MMPI Lie scale (Hathaway & McKinley, 1951): Since two of the standard 15 items appeared inappropriate for our college Ss, only 13 dichotomous-response items were used.⁴

7. Cattell's (1957b) Covert Anxiety scale included the first 20 "subtle" trichotomous-response (true-false) items of the IPAT Anxiety scale.

8. Cattell's (1957b) Overt Anxiety scale included the last "overt, symptomatic" items of the IPAT Anxiety scale.

9. Cattell's (1956b, p. 7) Difference score was the score for Overt Anxiety minus the score for Covert

Anxiety. Cattell suggests that the difference between the Covert and Overt scores "is intended to give a measure of the degree to which the patient is or is not conscious of his anxiety, and perhaps also of any wish consciously to emphasize it."

10. Cattell's Total Anxiety score was the sum of the Covert Anxiety and Overt Anxiety scores. Since these four scores derived from the IPAT Anxiety scale are not all experimentally independent, the Covert Anxiety and Overt Anxiety scores (which are independent) were included in the analysis of the data from one second group of Ss (Group A), while the Difference and Total Anxiety scores (which also are independent) were used in the analysis of the data from the second group of Ss (Group B).

11. Cattell's Neuroticism scale contained the 34 trichotomous-response items scored in the IPAT Neurotic Personality Factor questionnaire (Cattell, 1955). The items were reversed in scoring (high score representing high neuroticism) for consistency in scoring direction with other similar scales included in this study.

12. Cattell's (1956b) Motivational Distortion scale contains the remaining six trichotomous-response items in the IPAT NPF inventory (Cattell, 1955). Cattell's description of the construction of this scale suggests that it is measuring a factor similar to that assessed by the MMPI Lie scale.

In the above list, Variables 3, 4, 6, and 12 were included as marker variables for possible contaminating factors, while Variables 1, 2, 5, 7, 8, 9, 10, and 11 provided the test of the single Emotionality factor versus the double Anxiety-Neuroticism factor hypotheses.

Subjects. The scales were administered during a regular class session to 11 sections of introductory psychology, with a total of 425 Ss (261 men and 164 women) being tested. The Ss are primarily freshman and sophomore students in the School of the Liberal Arts with smaller proportions of juniors and seniors and of students enrolled in the School of Business Administration and the School of Engineering and Mines. After scoring, the answer sheets were alphabetized by Ss' names and the first 118 male Ss and the first 82 female Ss selected as Group A. In a similar manner the next 118 male and 82 female Ss were selected as Group B. The mean age of the 400 Ss was 20.8 years, with a standard deviation of 4.5 years. As noted above, the only procedural difference in analyzing the data from each group was that Cattell's Covert Anxiety and Overt Anxiety were included as score variables for Group A, while Cattell's Difference and Total Anxiety scores were used in the analysis of the data from Group B. Ss' sex (male scored 1, female scored 0) was included in each analysis as the eleventh variable.

Method. The 10 questionnaire variables within each group of Ss were intercorrelated by the usual product-moment method, and these variables were correlated with the sex dichotomy by the point biserial formula. Each matrix of 55 intercorrelations was factor analyzed by the complete centroid method, and four factors were extracted. Each of the succes-

¹ The MMPI booklet numbers can be found in Windle (1955).

² The MMPI booklet numbers and direction of scoring of these items are: 5, 29, 41, 44, 47, 72, 76, 108, 114, 159, 189, and 236 scored True; and 2, 3, 9, 46, 51, 68, 103, 175, 178, and 208 scored False.

³ The MMPI booklet numbers and direction of scoring of these items are: 54, 169, and 257 scored True; 40, 42, 138, 148, 156, 171, 218, 245, 247, 252, 267, 269, 286, and 383 scored False.

⁴ The MMPI booklet numbers are: 15, 30, 45, 60, 75, 90, 105, 120, 150, 165, 195, 225, and 285 scored False.

TABLE 1

ROTATED ORTHOGONAL FACTOR LOADINGS (DECIMAL POINTS OMITTED) OF INVENTORY SCORES

Scales	Group A				Group B			
	Em	F	Sx	h ²	Em	F	Sx	h ²
Taylor Manifest Anxiety Scale	83	-19	21	77	82	-32	12	79
Winne Neuroticism (Abbreviated)	61	-05	05	37	62	-35	-05	51
Edwards Soc. Desir. (Abbreviated)	-57	39	17	51	-54	46	14	53
Eysenck MPI Extraversion scale	-36	-04	-08	14	-46	-05	-03	22
Eysenck MPI Neuroticism scale	75	-33	12	68	67	-50	18	73
MMPI Lie scale (Abbreviated)	-07	66	05	44	-08	67	02	45
Cattell Covert Anx. subscale	50	-64	00	66				
Cattell Overt Anx. subscale	60	-37	40	66				
Cattell Difference score					47	09	37	37
Cattell Total Anxiety scale					61	-66	21	85
Cattell NPF Neuroticism scale	34	-09	66	56	42	-37	55	61
Cattell NPF Motiv. Dist.	-05	62	-12	40	-05	64	05	41
Sex (Female)	-02	03	56	31	-04	08	54	30

sive factors was tested for statistical significance using both Tucker's and Humphrey's criteria (Fruchter, 1954, pp. 77-80). The significant factors in each matrix were analytically rotated to orthogonal simple structure using the normalized varimax method (Kaiser, 1958).⁵

Results

The first four centroid factors from Groups A and B accounted for 53.2 and 54.8% of the variance of the 11 variables within each group of Ss, with the Group A factors accounting for 34.6, 9.1, 6.3, and 3.2 of the variance, while the similar percentages from Group B were 38.6, 8.7, 5.2, and 2.3. The application of Humphrey's rule in testing the significance of the factors indicated that the first three factors from Group A and the first two factors from Group B were significant at the .05 level or better. Using Tucker's procedure the first three factors were significant within both groups. Consequently, only three factors were retained for the normalized varimax rotation, with the fourth factor being discarded as not being significant under either criteria.⁶

⁵ The author's appreciation is extended to Gary Lotto and William B. Kehl of the University's Computing and Data Processing Center for providing assistance and facilities in the factor analytic phase of this study.

⁶ Tables giving the intercorrelations among the variables within each group of Ss, the tests of significance of the unrotated centroid factors, and centroid loadings and transformation matrices have been deposited with the American Documentation

The loadings of each variable resulting from the normalized varimax rotation procedure are given in Table 1. For convenience and consistency, the signs of the loadings for the Sex variable were reversed in preparing Table 1 as is indicated in the table. The three factors are almost identical between both Groups A and B, with indices of r^2 similarity (sum of the cross-products of loadings of similar factors divided by the geometric mean of the sums of squares of the loadings) for the nine variables common to both Groups A and B being .99, .94, and .96 for rotated Factors Em, F, and Sx. The only evident sampling difference between the independent groups of Ss is that the loadings of the scores on Factor F tend to be higher within Group B than they are within Group A. The relative rank orders of the Factor F loadings are almost identical within both groups ($\rho = .92$), but F accounts for only 15% of the interscale variance within Group A, while it accounts for 19% within Group B.

Factor Em is an Emotionality factor with zero loadings only on the MMPI Lie, Cattell Motivational Distortion, and Sex variables. Taylor's MAS is the best single measure of Factor Em, with Eysenck's Neuroticism, Winne's Neuroticism, and Cattell's Total Anxiety scores also having large loadings on this factor. This Emotionality factor, as de-

finer by this battery of inventories, permeates both the Anxiety and Neuroticism scales equally strongly and shows a moderate negative loading ($-.36$ and $-.46$) on the Extraversion scale. Factor F is a Falsification factor with high loadings on the MMPI Lie and Motivational Distortion scales. Since neither of these two variables was loaded on the other two factors, they can be used to define and identify this factor. Most of the scales with high Emotionality loadings also have moderate negative loadings on this F factor with Cattell's Covert Anxiety and Total Anxiety scales showing the largest loadings. Eysenck's Extraversion scale and the Sex variable have zero loadings on Factor F. Factor Sx appears to be a pure Sex factor with the reversed direction of scoring of this variable identifying the positive pole as the "femininity" direction. Both Cattell's Overt Anxiety and Neuroticism scales have large positive loadings on the Sx factor as does Cattell's Difference score (which is probably attributable to the inclusion of Covert Anxiety in the Difference score).

Considering the factor composition of the inventory scores included in this analysis, the Social Desirability, Eysenck Neuroticism, Covert Anxiety, Total Anxiety, and probably the Manifest Anxiety and Winne Neuroticism scales are factorially complex by being loaded on both Factors Em and F. Overt Anxiety and probably Cattell's Neuroticism scales are loaded with all three factors, while Cattell's Difference score is loaded with both Factors Em and Sx. Eysenck's Extraversion scale shows a moderate loading on Factor Em, and the MMPI Lie and Motivational Distortion scales are the best measures of Factor F.

Discussion

The absence of a significant unrotated fourth factor within either matrix of inventory intercorrelations and the clear identification of the analytically rotated three significant factors as Emotionality, Falsification, and Sex factors indicate that the Anxiety and Neuroticism dimensions measured by the scales developed by different authors and different methodologies can be assumed to be operationally identical within the limits of

this study. This is not to deny that some of these scales may be also measuring factors that had no opportunity to appear in these analyses, but the communalities of several scales, notably Taylor's MAS, Eysenck's Neuroticism, Cattell's Covert Anxiety, Overt Anxiety, Total Anxiety, and Difference scores, are close to what might be assumed to be the reliabilities of these scales, leaving little reliable scale variance to be attributed to hypothetical factors not appearing in these analyses.

The Emotionality factor defined by these inventories includes an introversion component as has been suggested by Eysenck (1957). His further hypothesis that this component results from the confounding of an orthogonal Extraversion-Introversion (EI) factor with Em could not be tested with the present data since only one "pure" marker variable for the hypothetical EI factor had been included in the analyses. Inclusion of additional measures of EI might permit a separation of Emotionality and Extraversion factors that may have been confounded in the present analyses and allow the assessment of the relative proportions of each factor in these Anxiety and Neuroticism scales. The presence of a Falsification or Honesty factor in these inventory scores, as was suggested by Edwards (1957) among others, was also strongly confirmed, but it is evident that Edwards' Social Desirability Scale is not the best measure of this factor. The factor loadings of the abbreviated Social Desirability Scale appear quite similar to those of the Anxiety and Neuroticism scales and indicate that the SDS scores are better measures of Factor Em than they are of Factor F. Although Eysenck (1956a) was moderately successful in developing an Emotionality scale that is independent of Extraversion (Bendig, 1957), his scale is also negatively loaded with the Falsification factor.

The study reported below is a further extension of the first study with additional marker variables for the Em and TH factors. It was hoped that these additional scales would clarify the relation of the scales to the Em, EI, F, and Sx factors and would again test for separate Anxiety and Neuroticism factors.

STUDY II

Procedure

Scales. Eleven independent personality scales were included, along with standard instructions, in a booklet to be distributed to the Ss with separate IBM answer sheets being used to collect the responses to the items.

Maudsley Personality Inventory (MPI) scores: The most recent form of the MPI, described in detail by Jensen (1958), consists of 80 trichotomous-response (Yes-?-No) items including 24 Extraversion, 24 Neuroticism, 20 Lie, and 12 Buffer items. Eysenck (1956a) describes the 12 Buffer items as 7 nonscored Extraversion and 5 nonscored Neuroticism items to be used for research purposes. He also indicates which of the 24 items on either the E or the N scales can be split into subscales containing 12 items each for establishing the split-half reliability of the E and N scales. Much of the content of the Lie items is similar to that in other Lie scales such as the MMPI scale described below. Although the original MPI uses the trichotomous-response method, it has been shown that the reliability and intercorrelation of the scales are not affected if a simpler dichotomous-response (Yes-No) method is used (Bendig, 1959a).

For purposes of the present research, the items of the MPI, using the dichotomous-response format, were grouped into six independent scales to obtain separate scores. MPI-E1 and MPI-E2 each contained 12 of the original 24 Extraversion items with the assignment of items to subscales following the split suggested by Eysenck (1956a). MPI-E3 included the 7 Buffer MPI Extraversion items. MPI-N1 contained the 24 items in the original Neuroticism scale, while MPI-N2 was scored for the 5 Buffer MPI Neuroticism items. MPI-L scores were obtained by scoring 10 of the 20 MPI Lie items, omitting 10 Lie items that either overlapped in item content some of the other Lie scales used in this study or appeared not to be items which would be relevant for our Ss.⁷

Some of the scales included in the previous study were repeated in this analysis. Taylor's 50-item Manifest Anxiety Scale (MAS) was included along with the abbreviated 22-item version of Winne's Neuroticism (Winne N) scale and the 13-item form of the MMPI Lie scale. The 34-item Neuroticism (NPF-N) and 6-item Motivational Distortion (NPF-MD) scales included in Cattell's Neurotic Person-

ality Factor Questionnaire were also repeated in this study. As in the previous study, the NPF Neuroticism items were reversed in scoring compared to the item key given in the NPF manual so that high scores here represent high Neuroticism.

In summary, the MPI-N1, MAS, Winne N, MMPI Lie, NPF-N, and NPF-MD scales are identical with those used previously, while the MPI-E1 and MPI-E2 combined are equivalent to the previous Extraversion scale. MPI-E3, MPI-N2, and MPI-L were used for the first time in this study.

Subjects. The scales were administered to a total of 263 students (197 men and 66 women) enrolled in eight sections of a one-semester course in introductory psychology (Group C). The Ss were primarily freshman and sophomore students and are similar to those used in the previous study with the exception that a smaller percentage of women Ss was enrolled during this (spring) semester. The scales were administered in class during the second week of the semester by each class instructor, and this group of Ss represents the total enrollment in the course during this semester with the exception of a few Ss who were absent from class the day that the scales were administered.

Method. The 11 scale scores were intercorrelated by the usual product-moment method, and each score was correlated with the sex dichotomy by the point biserial formula. Six factors were extracted from the 12-variable matrix of correlations by the complete centroid method and each factor was tested for statistical significance using both Humphrey's and Tucker's criterion. The significant factors were rotated to orthogonal simple structure using the analytic normalized varimax criterion.

Results

Using Humphrey's test of the significance of factors, the first three centroid factors were significant at the .01 level and the fourth factor at the .05 level ($z = 2.20$), while the fifth and sixth factors did not approach significance ($z = .82$ and $.53$). The first five factors were significant using Tucker's phi criterion, while the sixth factor again was not significant. Including the fifth factor in the rotation process resulted in a fifth rotated factor with all loadings less than .30 (the largest loadings were $-.28$ for NPF-N and $.22$ for MMPI Lie). Consequently, the fifth factor was omitted and only the first four factors included in the normalized varimax rotation procedure. The four unrotated centroid factors accounted for 53.8% of the total variance with the omitted fifth factor accounting for only an additional 2.0%. The residual correlations after the extraction of four factors ranged in absolute

⁷ The content of the MPI items and item scoring direction can be found in Jensen (1958, pp. 316-318). The numbers of the items included in this study are: MPI-E1, 1, 3, 6, 13, 14, 20, 31, 56, 57, 66, 69, 75; MPI-E2, 2, 9, 23, 37, 41, 44, 46, 47, 51, 61, 73, 79; MPI-E3, 7, 21, 29, 34, 60, 70, 80 (all scored for Yes responses); MPI-N2, 22, 30, 53, 74, 78 (all scored Yes); and MPI-L, 5, 12, 19, 33, 36, 40, 43, 49, 58, 62. The MPI-N1 scale included the 24 items listed by Jensen for the MPI-N scale. Lie items 8, 15, 24, 28, 45, 52, 55, 67, 72, and 77 were omitted from the form given to the Ss.

TABLE 2

ROTATED ORTHOGONAL FACTOR LOADINGS
(DECIMAL POINTS OMITTED) OF
INVENTORY SCORES (Group C)

Scales	Em	EI	F	Sx	h ²
Eysenck MPI-N1	78	-07	-25	14	70
Eysenck MPI-N2	64	-01	-13	10	43
Taylor MAS	85	-23	-12	07	80
Winne N	69	-23	-04	04	53
Cattell NPF-N	46	04	-13	52	50
Eysenck MPI-E1	-22	78	01	-10	66
Eysenck MPI-E2	-10	79	-02	11	64
Eysenck MPI-E3	-05	79	08	02	64
Eysenck MPI-L	02	04	62	06	39
MMPI Lie	-27	-10	56	10	40
Cattell MPF-MD	-22	11	55	-11	38
Sex (Female)	05	01	09	59	36

values from .00 to .08, with the median being .015.

The rotated factor loadings show a good approximation to simple structure, with 45.8% of the loadings having absolute magnitudes of .10 or less, 60.4% being .20 or less, and 27.1% of the loadings being significant (.30 or greater). Only NPF-N had significant loadings on two factors with the other 11 variables having loadings above .30 on one factor. The factor loadings and communalities of the scales can be found in Table 2. The first factor is obviously the Emotionality (Em) factor with high loadings on all of the Anxiety and Neuroticism scales, while the orthogonal second factor is an Extraversion-Introversion (EI) dimension which loads highly on the three Extraversion subscales. The third rotated factor again is a Falsification (F) factor with high loadings on the three Lie scales and the fourth factor is a Sex (Sx) factor that loads only on the NPF-N and Sex (female) variables. MPI-N1 has a small negative loading on F, while the MAS and Winne N variables have small negative loadings on EI. MMPI Lie and NPF-MD scales have small negative loadings on Em in addition to their large loadings on F with MPI-L loading only on F. NPF-N loads on both Em and Sx with its loading on Em (.46) being slightly smaller than its loading on Sx (.52).

Discussion

The results of this factor analysis have confirmed and clarified the findings of the previous analyses. The presence of an orthogonal EI factor is confirmed and has been separated from the Em factor, while the previous analyses confounded these two factors. Again only one major Em factor was found to load the Anxiety and Neuroticism scales, which confirms the previous evidence that Anxiety and Neuroticism factors cannot be demonstrated as separate factors in factor analyses of questionnaire data. The identification of the factor was strengthened and more clearly separated from the EI factor.

Using the squares of the factor loadings in Table 2 as percentages of scale variance, we can quantitatively estimate factor components in the scales included in this research. Of the Anxiety and Neuroticism scales, the MAS is the most valid measure of Em (72%), but is contaminated by a small loading on EI (5%). Eysenck's MPI-N1 score is also a good measure of Em (61%) and is free of EI, but is slightly loaded on F (6%). Winne's Neuroticism scale shows a lower weighting on Em (48%) and, like the MAS, has an EI component (5%). Cattell's NPF Neuroticism scale is a better measure of Sex differences (Sx) than it is of Emotionality (21% vs. 27%) which agrees with the previous analyses. The greater saturation of the MAS score with Em may be due to the greater length of this scale (50 items) compared to the other scales, assuming that a longer scale would have a greater proportion of reliable variance and, possibly, a larger proportion of common factor variance.

Approximately 62% of Eysenck's Extraversion scale variance is accounted for by the EI factor and 3% is attributable to Em. All three of the Lie scales (MPI-L, MMPI Lie, and NPF-MD) load moderately on the F factor (38, 31, and 30%), but MMPI Lie and NPF-MD also have Em components (7 and 5%). MPI-L appears to be the purest measure of this F factor with negligible loadings on Em, EI, and Sx. It is interesting to note that the three scales composed of MMPI items (MAS, Winne, and MMPI Lie) all have both Em and EI components.

In using these scales for further factor analytic studies, MAS, MPI-N1, and MPI-N2 appear to be the best "markers" for the Em factor and Cattell's and Winne's Neuroticism scales can be eliminated as being too contaminated by extraneous factor variance. However, the Winne scale might be improved and purified through factor analysis. The MPI Extraversion scores can be used to define the EI factor, but should be improved to remove the F component. MPI-L can be used as a measure of F, but its F saturation should be increased, possibly by lengthening the scale, and at least one other F marker scale will have to be developed that is free of the Em factor.

SUMMARY

Various inventories measuring Anxiety, Neuroticism, Extraversion, and Falsification variables were administered to two groups of college students, each group containing 200 Ss, and replicated factor analyses of the intercorrelations among 11 scores within each group were performed. Three significant factors were found within each matrix of correlations and the factors were rotated to orthogonal simple structure by the normalized varimax method. The rotated factors were identified as Emotionality, Falsification, and Sex with all of the Anxiety and Neuroticism scores having appreciable loadings on both of the first two factors. It was suggested that Anxiety and Neuroticism are both manifestations of a more general Emotionality factor and are not separate dimensions within commonly used inventories, and that the analyses may have confounded two independent factor dimensions, Emotionality and Extraversion-Introversion, within the Emotionality factor found in these two samples.

To test the above findings, a second study was conducted in which 11 Anxiety, Neuroticism, Extraversion, and Lie scales overlapping with the first battery were administered to 263 students and a centroid factor analysis performed on the matrix of scale intercorrelations. Four factors were found which, after rotation to orthogonal simple structure, were clearly identifiable as Emotionality (Em), Extraversion-Introversion (EI), Falsification

(F), and Sex (Sx). Separate factors of Anxiety and Neuroticism again did not appear in the analysis, and the separation of Emotionality and Extraversion-Introversion as independent factors was confirmed.

REFERENCES

- BENDIG, A. W. Extraversion, neuroticism, and manifest anxiety. *J. consult. Psychol.*, 1957, 21, 398.
- BENDIG, A. W. Identification of item factor patterns within the Manifest Anxiety Scale. *J. consult. Psychol.*, 1958, 22, 158.
- BENDIG, A. W. Score reliability of dichotomous and trichotomous item responses on the Maudsley Personality Inventory. *J. consult. Psychol.*, 1959, 23, 181. (a)
- BENDIG, A. W. "Social desirability" and "anxiety" variables in Cattell's IPAT Anxiety scale. *J. consult. Psychol.*, 1959, 23, 377. (b)
- CATTELL, R. B. *The neurotic personality factor questionnaire*. Champaign, Ill.: Institute for Personality and Ability Testing, 1955.
- CATTELL, R. B. A shortened "basic English" version (Form C) of the 16 P-F questionnaire. *J. soc. Psychol.*, 1956, 44, 257-278.
- CATTELL, R. B. The conceptual and test distinction of neuroticism and anxiety. *J. clin. Psychol.*, 1957, 13, 221-233. (a)
- CATTELL, R. B. *Handbook for the IPAT Anxiety scale*. Champaign, Ill.: Institute for Personality and Ability Testing, 1957. (b)
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- EYSENCK, H. J. The questionnaire measurement of neuroticism and extraversion. *Riv. Psicol.*, 1956, 50, 113-140. (a)
- EYSENCK, H. J. Reminiscence, drive, and personality theory. *J. abnorm. soc. Psychol.*, 1956, 53, 328-333. (b)
- EYSENCK, H. G. *The dynamics of anxiety and hysteria*. London: Praeger, 1957.
- FREY, A. H., & BECKER, W. C. Some personality correlates of subjects who fail to appear for experimental appointments. *J. consult. Psychol.*, 1958, 22, 164.
- FRUCHTER, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.
- HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory*. (Rev. ed.) New York: Psychological Corp., 1951.
- HOLTZMAN, W. H., CALVIN, A. D., & BITTERMAN, M. E. New evidence for the validity of Taylor's Manifest Anxiety Scale. *J. abnorm. soc. Psychol.*, 1952, 47, 853-854.
- JENSEN, A. R. The Maudsley Personality Inventory. *Acta psychol.*, Hague, 1958, 14, 314-325.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.

- McCREARY, JOYCE, & BENDIG, A. W. Comparison of two forms of the Manifest Anxiety Scale. *J. consult. Psychol.*, 1954, 18, 206.
- O'CONNOR, J. P., LORR, M., & STAFFORD, J. W. Some patterns of manifest anxiety. *J. clin. Psychol.*, 1956, 12, 160-163.
- SPENCE, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, 13, 131-141.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- WINDLE, C. The relationships among five MMPI "anxiety" indices. *J. consult. Psychol.*, 1955, 19, 61-63.
- WINNE, J. F. A scale of neuroticism: An adaptation of the Minnesota Multiphasic Personality Inventory. *J. clin. Psychol.*, 1951, 7, 117-122.

(Received April 20, 1959)

PERCEPTION OF MATERNAL CHILD REARING ATTITUDES IN SCHIZOPHRENICS

ALFRED B. HEILBRUN, JR.

State University of Iowa

Theories regarding the conditions antecedent to the development of schizophrenia still remain speculative in character. The points of view fall mainly into three categories: causative factors are of a genetic or constitutional nature; causative factors are of a psychological character linked to some maladjustment in interpersonal relationships; causative factors represent a combination of physical and psychological agents. Many of those who have attempted to account for schizophrenia in terms of psychological factors have stressed the importance of faulty mother-child relationships in the etiology of this disorder (Arieti, 1955; Fromm-Reichmann, 1948; Rosen, 1947). One line of research bearing upon the mother-child hypothesis has utilized an objective measure of the child rearing attitudes of the schizophrenics' mothers, comparing them to the mothers of normals.

Mark (1953) reported that 67 of the 139 items in an attitude survey reliably discriminated between the mothers of male schizophrenics and the mothers of normal males with the schizophrenics' mothers tending to be more "restrictive" and either "excessively devoted or detached." Using Shoben's (1949) Parent-Child Attitude Survey, Freeman and Grayson (1955) found significant or suggestive differences on two out of three subscales when the mothers of schizophrenics and normals were compared. Tendencies to be "possessive" or "ignoring" in attitude were shown by the mothers of schizophrenics. However, a study by Zuckerman, Oltean, and Monashkin (1958a), which compared the responses of these two types of mothers on the Parent Attitude Research Instrument (Schaefer & Bell, 1958), found only one significant difference between them on 20 scales, a finding which can be readily attributed to chance.

The apparent logic behind these studies and numerous others (see Freeman & Grayson, 1955) in which a variety of assessment techniques have been used to evaluate the maternal attitudes of schizophrenics' mothers is that the best way to detect and clarify possible malignant factors in the mother-schizophrenic child relationship is to directly evaluate the mother. However, it can be argued that it is the child's *perceptions* of the mother's characteristics and attitudes which are the important cues mediating his behavior in the mother-child relationship. It would follow that the study of the possible genesis of schizophrenia in disruptive mother-child interactions should proceed by way of the schizophrenic child's perceptual schema of the mother and of himself in relation to her. An earlier study (Singer, 1954) followed a similar approach by using Thematic Apperception Test stories to evaluate schizophrenics' perception of parental characteristics and intrafamilial relationships. Singer concluded that schizophrenics see parental figures as cold or rejecting and depict the relationships with parents as being relatively isolated.

The present study represents an attempt to measure objectively schizophrenic perceptions of mothers' child rearing attitudes and to compare these perceived attitudes with those of normals. More specifically, the hypothesis that schizophrenics would perceive the mother-child relationship as poorer than would normals was tested by predicting that schizophrenics would attribute more pathogenic attitudes to their mothers than would normals.

METHOD

Subjects

Twenty-six females, admitted in 1957 to a typical large midwestern mental hospital and who carried un-

contested staff diagnoses of schizophrenia, were individually administered the Parent Attitude Research Instrument (PARI). There was no attempt made to provide more homogeneous classes of schizophrenics by further categorizing them into the standard subgroups of paranoid, catatonic, etc. One reason for this was the restricted number of schizophrenics available for the study since inclusion demanded the ability and cooperation of *both* the schizophrenic and her mother. The second reason was the lack of any evidence that membership in any subclass was differentially related to the experimental task. A control group of 27 normal females (i.e., answered "No" when asked if they had ever utilized any mental health service) were personally contacted or contacted by mail and were willing to complete the PARI.

The schizophrenic and normal groups did not differ significantly in mean age (schizophrenics, 36.6 years; normals, 33.4 years) or mean educational level (schizophrenics, 11.3 years; normals, 12.1 years). Hospitalization was only partially controlled (26 out of 27 of the schizophrenics were hospitalized at the time of testing while 10 of 26 normals were patients on medical wards of a general hospital) but there is no *a priori* basis for assuming a *systematic* relationship between the hospitalization factor and the test performance obtained in the present study. The social-economic status variable was not controlled in this study except to the extent that having fairly equivalent mean educational levels for the daughter groups accomplishes this. Since social-economic status has been shown to be related to child rearing attitudes (though not to the *perception* of such attitudes), it seemed necessary to make a preliminary evaluation of whether the mothers of the two daughter groups differed attitudinally. The mothers were also given the PARI and an analysis of variance demonstrated no differences ($F < 1.00$) in expressed child rearing attitudes between the mothers of the schizophrenic and normal daughters used in this study.

Test

The attitude survey used in the present investigation was the Parent Attitude Research Instrument (PARI) developed by Schaefer and Bell (1958). It includes 115 statements devised to measure maternal attitudes judged to be related to personality development in children. The total of 115 statements consists of 23 sets of 5 statements each, with each set intended to measure a category of maternal attitudes. Descriptive scale names are given in Table 1. The test taker has the option of endorsing "strongly agree," "mildly agree," "mildly disagree," or "strongly disagree" for each statement. A "strongly agree" response is given the highest score on the PARI and "strongly disagree" the lowest; therefore, a high score always represents agreement with the attitude in question. Further, the attitudes being measured by 20 of the scales taken in their extreme are typically assumed to be deviant when characteristic of

the mother. Accordingly, higher scores for these 20 scales indicate both greater agreement and a greater potential disruptive influence on the child's development. The remaining three "rapport" scales (Nos. 1, 14, 21 on Table 1) were included by the test constructors so that normals would have a greater opportunity to agree with statements on the questionnaire. Scores on these three scales are interpreted in opposite fashion so that higher scores indicate more *adjustive* attitudes.

Procedure

The schizophrenic daughters were individually administered the PARI under special instructions to "Answer it as you think your mother would." The administration was conducted by some member of the state hospital staff who was previously known to the patient. The staff member was requested to be sure the nature of the task was initially clear to the patient and to restate the directions periodically if it seemed necessary. The 26 schizophrenics included in this study represent about 29% of the 90 female patients with appropriate diagnoses admitted during the data collection period.

The 10 normal daughters obtained from general medical wards were individually administered the PARI by a psychologist under the same conditions as the schizophrenics. The remaining 17 normal daughters were contacted by mail and were strongly urged to complete the task without prior discussion with their mothers concerning it. The names and addresses of these 17 daughters were obtained from their mothers who were initially tested. As far as it is known, the daughters were not acquainted with one another. The mothers, in turn, were obtained as Ss from the Iowa City Employment Agency and local civic groups. The total sample of 27 daughters represents about 42% of the 65 requested to participate in the study.

The printed task instructions were identical for both groups, whether seen personally or contacted by letter. Further, the impersonal research nature of the task was stressed for all daughter Ss to decrease test response manipulation associated with public description of their mothers.

Scoring of the predictions proceeded just as though the mother had completed the PARI under standard conditions with a score of 4 assigned to a "strongly agree" prediction for a statement, 3 to a "mildly agree," 2 to a "mildly disagree," and 1 to "strongly disagree." These statement scores were then summed over the five statements within each scale. The scale prediction scores on the PARI define perceived maternal attitudes in this study.

RESULTS

The data presented in Table 1 were examined initially by a Type I analysis of variance (Lindquist, 1953) in which the main effect for diagnosis was a between-S effect

TABLE 1

COMPARISONS OF MEAN MOTHERS' ATTITUDE SCALE SCORES AS PREDICTED BY SCHIZOPHRENIC AND NORMAL DAUGHTERS

Scale Number and Description	Primary Factor Loading ^a	Schizophrenic (N = 26)		Normal (N = 27)		t value
		Mean	SD	Mean	SD	
2 Fostering Dependency	A	14.10	4.00	11.25	3.30	2.85**
3 Seclusion of Mother	A	14.55	3.05	12.45	2.30	2.80**
4 Breaking the Will	A	13.60	3.10	11.25	3.25	2.76**
5 Martyrdom	A	13.75	3.85	10.80	3.60	2.81**
6 Fear of Harming Baby	A	14.80	3.40	14.00	3.30	.89
10 Excluding Outside Influences	A	13.00	2.05	11.45	2.35	2.58**
11 Deification	A	15.35	3.30	13.40	3.30	2.17*
12 Suppression of Aggression	A	13.50	2.75	11.15	1.90	3.62**
15 Approval of Activity	A	14.65	2.65	13.60	3.00	1.31
16 Avoidance of Communication	A	12.90	3.85	10.60	3.10	2.42*
17 Inconsiderateness of Husband	A	13.85	3.55	12.60	4.00	1.19
18 Suppression of Sex	A	13.25	3.50	10.65	3.60	2.74**
19 Ascendancy of Mother	A	14.85	3.50	11.80	2.80	3.59**
20 Intrusiveness	A	12.85	3.65	10.60	3.30	2.37*
22 Acceleration of Development	A	15.35	2.60	12.45	3.85	3.05**
23 Dependency of Mother	A	14.20	2.60	12.85	3.40	1.59
7 Marital Conflict	B	14.80	2.75	15.80	3.05	1.25
9 Irritability	B	13.25	2.30	14.80	3.10	2.07*
13 Rejecting Homemaking Role	B	12.00	2.80	11.60	3.95	.42
1 Encouraging Verbalization	C	16.55	2.30	16.75	2.55	.29
14 Equalitarianism	C	16.50	2.60	16.40	2.75	.13
21 Comradeship and Sharing	C	17.85	1.70	18.35	1.45	1.11
8 Strictness		14.45	3.45	13.75	2.95	.78

^a Taken from Zuckerman, Ribback, Monashkin, & Norton, 1958.

* Significant beyond .05 level.

** Significant beyond .01 level.

and the main effect for scales was based upon within-S comparisons. The *F* value for the diagnosis \times scales interaction was significant beyond the .001 level of confidence (mean square for interaction = 1.09; mean square for within treatments = .28; *F* = 3.89 for 22, 1122 *df*). Since the diagnosis \times scales interaction effect was significant, no evaluation of the main effects of diagnosis or scales was made. Rather, interest was focused upon the simple effects of diagnosis on each of the individual scales (Lindquist, 1953, p. 211). Table 1 presents the comparisons of predicted attitude scores for the individual PARI scales. Inspection of this table shows that in 13 of the 23 scale comparisons, the schizophrenics' prediction scores differed from those of the normals. Of the 13 comparisons providing significant differences, 12 show the schizophrenic daughters ascribing attitudes to their mothers which are more contrary to usually

approved child rearing opinions than are those ascribed to their mothers by normals. The exception was the Irritability Scale where normal daughters predicted that their mothers would score higher than did the schizophrenics.

DISCUSSION

When schizophrenic daughters were asked to predict how they thought their mothers would perform on a survey of child rearing attitudes and these predictions were compared with those provided by normal daughters, marked differences in these perceptions were obtained. The major trend was toward the schizophrenic's perception of her mother as endorsing less approved (and presumably more pathogenic) child rearing attitudes when compared to the normal's perception of her mother. This trend was observed for 12 of the 23 PARI attitude scales. Normal daughters

ters predicted that their mothers would endorse a more pathogenic attitude on only one scale and on the ten remaining scales no significant differences were obtained.

These findings lend support to the notion that the schizophrenic daughter attributes more pathogenic child rearing attitudes to her mother than does the normal daughter. It may be recalled that the mothers of these daughter groups did not differ significantly in their expressed child rearing attitudes on the same attitude scale used by the daughters, a finding which is consistent with the study of Zuckerman et al. (1958a). However, since this study is concerned with perceptual differences between schizophrenics and normals and not with accuracy of perception, this discrepancy between perception and endorsement is not of immediate concern.

Factor analytic studies with the PARI (Schaefer & Bell, 1955; Zuckerman, Ribback, Monashkin, & Norton, 1958b) have found three stable factors which reportedly account for most of estimated communality variance on this instrument. Factor A has been defined as an "authoritarian-control" cluster; Factor B, "hostility-rejection"; and Factor C, "democratic attitudes." An analysis of the findings in the present study shows that 12 of the 16 scales with highest loadings on Factor A (see Table 1) were characterized by a significantly greater perceived pathology in maternal attitudes on the part of schizophrenics. On the other hand, Factor B included scales where schizophrenic-normal comparisons produced a far different picture. On one of the scales (9), normals predicted a reliably higher score for their mothers; the two others (7, 13) produced no differences between the diagnostic groups. Factor C included three scales none of which produced a difference between diagnostic groups. These findings suggest that it is in the "authoritarian-control" aspects of mother-child relationships where schizophrenics perceive their mothers as deviating more from accepted child rearing practices than do normals. Further the "hostility-rejection" dimension either does not distinguish schizophrenics' and normals' perceptions of their mothers or normals perceive their mothers as deviating more from standard practices in this regard. Finally, there is

no indication of a difference between perceptions of these diagnostic groups with regard to "democratic attitudes" of the mothers.

Thus when the child rearing attitude patterns of the mothers of schizophrenic and normal daughters are examined through the daughters' perception of them and these perceptions are evaluated in rather molar conceptual units, it begins to appear that a major source of interpersonal difficulty in the mother-schizophrenic daughter relationship evolves from the perceived "authoritarian-control" tendencies of the mother.

It might be contended that the findings of the present study are not appropriate evidence regarding the etiological relationship between faulty mother-child relationships and the development of schizophrenia. The argument would be that the perceived child rearing attitudes of the adult schizophrenic are retrospective in nature and might vary markedly from those held as a child (or at least prior to the onset of schizophrenia). It is true that what appear to be systematic perceptual differences between schizophrenics and normals in this investigation can be speculatively related to schizophrenia in a number of ways of which causal significance is but one. In any case, definitive evidence is not to be found in the present study but rather from studies specifically designed to critically evaluate various alternative hypotheses.

It is also recognized that the results of the present study could be accounted for by differential test-taking habits of the schizophrenic and normal daughters rather than differential perceptual habits as previously discussed. For example, if these groups differed substantially along a social desirability response set continuum, one possibility would be that normals would attempt to depict their mothers in a more favorable light than the schizophrenics. Under such a condition, the direction of between-group differences on the PARI would undoubtedly be consistent with those found in this study. However, there are three reasons why accounting for the results in terms of differential social desirability response set tendencies appears less parsimonious than doing so in terms of perceptual differences:

1. The more impersonal research approach was stressed rather than an interest in any given individual's conception of her mother. This might be expected to lessen the likelihood that desirability of response would play a major role within either daughter group.

2. Previous studies comparing psychiatric and normal Ss (Gough, 1950; Klett, 1957; McKinley, Hathaway, & Meehl, 1948) indicate that these groups do not show differential characteristics with regard to desirability of test responses.

3. It would be difficult to account for finding systematic differences in predicted attitudes solely within the "authoritarian-control" cluster by presuming differential social desirability response sets. It would seem necessary to introduce a cumbersome triple interaction between social desirability, attitude cluster, and diagnosis to explain why 75% of the scales included in one cluster provided significant unidirectional differences between schizophrenics and normals because of the social desirability factor while not one scale outside of the cluster even approached that result.

SUMMARY

This study compared the perceptions of their mothers' child rearing attitudes on the part of schizophrenic and normal daughters. The two daughter groups, matched for age, education, and partially for hospitalization, were asked to predict how their mothers would respond to items on the Parent Attitude Research Instrument and these predictions operationally defined perceived maternal attitudes. The mothers of the schizophrenics and normals independently completed the PARI, and there were no differences in endorsed child rearing attitudes obtained between these mother groups.

Analysis of the predicted scale scores revealed an overall tendency for schizophrenic daughters to predict attitude scores for their mothers which reflected more variant child rearing practices than did the predicted attitude scores for the normals. These perceived attitude scale scores were further analyzed in terms of three primary factors obtained in previous factor analytic studies. This analysis provided rather convincing evidence that schizophrenics, more than normals, tend to perceive attitudes relating to the "authoritarian-control" dimension as more abnormal

in their mothers. However, perceptions of attitudes relating to "hostility-rejection" or to "democratic attitudes" provided no substantial basis for differentiation of these two classes of mothers.

REFERENCES

- ARIETI, S. *Interpretation of schizophrenia*. New York: Brunner, 1955.
- FREEMAN, R. V., & GRAYSON, H. M. Maternal attitudes in schizophrenia. *J. abnorm. soc. Psychol.*, 1955, 50, 45-52.
- FROMM-REICHMANN, FRIEDA. Notes on development of treatment of schizophrenics by psychoanalytic therapy. *Psychiatry*, 1948, 11, 263.
- GOUGH, H. G. The *F* minus *K* dissimulation index for the MMPI. *J. consult. Psychol.*, 1950, 14, 408-413.
- KLETT, C. J. The social desirability stereotype in a hospital population. *J. consult. Psychol.*, 1957, 21, 419-421.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Cambridge: Riverside, 1953.
- MARK, J. C. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.*, 1953, 48, 185-189.
- McKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The MMPI: VI. The *K* Scale. *J. consult. Psychol.*, 1948, 12, 20-31.
- ROSEN, J. N. The treatment of schizophrenic psychoses by direct analytic therapy. *Psychiat. Quart.*, 1947, 21, 3-37.
- SCHAEFER, E. S., & BELL, R. Q. Development of a parental attitude research instrument. *Child Development*, 1958, 29, 339-361.
- SCHAEFER, E. S., & BELL, R. Q. Parental Attitude Research Instrument (PARI). Normative data. Unpublished manuscript. National Institutes of Health, Bethesda, Md., 1955.
- SHOBEN, E. J., JR. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.*, 1949, 39, 101-148.
- SINGER, J. L. Projected familial attitudes as a function of socioeconomic status and psychopathology. *J. consult. Psychol.*, 1954, 18, 99-104.
- ZUCKERMAN, M., OLTEAN, MARY, & MONASHKIN, I. The parental attitudes of the mothers of schizophrenics. *J. consult. Psychol.*, 1958, 22, 307-310. (a)
- ZUCKERMAN, M., RIBBACK, BEATRICE B., MONASHKIN, I., & NORTON, J. A. Normative data and factor analysis on the Parental Attitude Research Instrument. *J. consult. Psychol.*, 1958, 22, 165-171. (b)

(Received April 22, 1959)

AMBIGUITY AND DISCRIMINATING POWER IN PERSONALITY INVENTORIES

WILLIAM E. BROEN, JR.

University of California, Los Angeles

Research and test construction procedures designed to improve personality inventories can be most fruitful if they are guided by a clear understanding of the rationale of personality inventories. However, disagreements concerning the beneficial or deleterious influence of personality inventory item ambiguity (Allport, 1937; Anastasi, 1954; Benton, 1935; Eisenberg, 1941; Elias, 1951; Freeman, 1955; Hutt, 1945; Landis & Katz, 1934; Meehl, 1945; Rotter, 1954; Vernon, 1953) indicate that, at least with regard to item ambiguity, this clarification of rationale has not been fully achieved. This paper is an attempt to arrive at some further clarification of personality inventory rationale through a discussion of item ambiguity and its relationships to discriminating power. The research and test construction implications of these relationships will also be discussed.

TWO TYPES OF AMBIGUITY

The classical position concerning the relation of item discriminating power and ambiguity is given by Allport in his listing of freedom from ambiguity as a requisite to the use of personality inventories. He states this position as follows: "The stimulus-situation is assumed to be identical for each subject, and his response is assumed to have a constant significance" (Allport, 1937, p. 449). As can be seen from the two halves of Allport's statement, item ambiguity can be divided into two types: variation in the subject's interpretation of what is meant by an item and variation in the extent to which different types of persons give similar responses to an item. These two kinds of ambiguity will be referred to as "interpretive ambiguity" and "response ambiguity," respectively.

Interpretive Ambiguity

The most widely held point of view concerning interpretive ambiguity is that freedom from interpretive ambiguity is necessary in order to achieve adequate discriminating power. This point of view was first reflected in the theoretical arguments of Allport (1937) and in the writings of earlier researchers who found interpretive ambiguity in inventory items and criticized personality inventories because of this ambiguity (Benton, 1935; Eisenberg 1941; Landis & Katz, 1934). The assumption that personality inventories must be free of interpretive ambiguity is also stated by Hutt (1945), and is seen in Taylor's (1953, p. 289) comments on the development of the Manifest Anxiety Scale where, without study of item discriminating power, the majority of the test items were rewritten "so as to minimize confusion and misinterpretation." This assumption is most recently reflected in current texts which list interpretive ambiguity as one of the primary problems of personality inventories (Anastasi, 1955, p. 558; Freeman, 1955, p. 499; Rotter, 1954, p. 257). Among the recent writers, this position is stressed most by Rotter (1954, p. 257), who lists similarity of subject interpretation of questionnaire items as one of the three basic assumptions of personality inventories.

Meehl (1945) was among the first to question the necessity for interpretively unambiguous personality inventory items. He takes the position that discriminating power is the prime requirement for inventory statements, and if discrimination is obtained through empirical item selection, the question of whether or not the discriminating items are interpre-

tively ambiguous is essentially an irrelevant question.

Vernon (1953) also takes the point of view that the question of item ambiguity is not itself a serious question for personality inventories. However, his rationale is different from Meehl's. Vernon (1953, p. 140) sees the basic influence of item ambiguity as harmful but feels that this negative influence is in some way ruled out by the cancelling effects of various interpretive choices over a series of test items. This cancelling effect is based on an assumption that the interpretive choices are randomly distributed among the groups being discriminated.

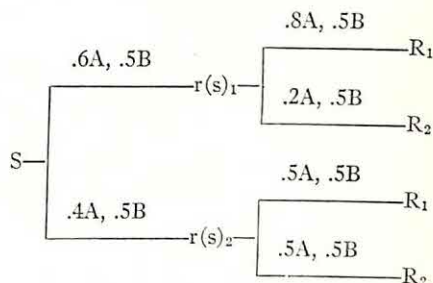
Elias (1951) goes even further and takes the point of view that interpretive ambiguity is useful and needed for maximal discrimination. His point of view is that the different individual interpretations of inventory statements may themselves reflect relevant personality differences and are thus sources of potentially discriminating variance. "An unstructured question permits greater leeway for personality projection; therefore, it should reflect personality tendencies more sensitively than more structured items" (p. 497).

If the primary task of personality inventories is to classify persons into relevant homogeneous groups, Meehl and Elias are correct in pointing out that amount of interpretive ambiguity is a relevant consideration only to the extent that interpretive ambiguity affects items discriminating power. Criticisms of personality inventories for interpretive ambiguity must then rest on the demonstration of a negative correlation between item interpretive ambiguity and discriminating power.

Partial research evidence on this point is given by Benton (1935), who found no correlation between interpretive ambiguity and discriminating power. These data would seem to contradict the position of those criticizing inventories because of interpretive ambiguity. Benton's data also suggest that there are no interpretive differences between "normal" and "abnormal" subjects, and thus also seem to negate Elias' position that different interpretations yield data relevant to personality categorization. However, the sample of items used in Benton's research is opened to question by Eisenberg's (1941) study where in-

terpretive variations were found to be discriminating. Thus, while suggesting rather clearly that it is not necessary to assume freedom from interpretive ambiguity in order to obtain discrimination, the available research evidence does not go much further in clarifying the relations between discriminating power and interpretive ambiguity. The presentation of a simplified ambiguity model may help in clarifying some of the relations involved.

An ambiguity model. Personality inventory item ambiguity can be most simply represented by a situation in which an inventory item (S) may lead to either of two mediating interpretations ($r(s)_1$, $r(s)_2$), both of which may then be responded to in terms of either Yes (R_1) or No (R_2). As an example, the various probabilities of mediating interpretations and responses for two equal N groups (A and B) being differentiated by an item can be represented as is illustrated below.



The above example assumes that 60% of Group A and 50% of Group B make the first mediating interpretation, and for the As making this interpretation 80% respond Yes (R_1). Multiplying interpretation probabilities by response probabilities we find that in this example 68% of the As and 50% of the Bs answer Yes to the item, and 32% of the As and 50% of the Bs respond No. Thus, if all who respond Yes are labeled as A, the statement correctly identifies 59%.

Trying out various combinations of interpretation and interpretation response probabilities, and studying the resultant discriminating power changes, indicate that the following can be used as a generalized formula for the discriminating power of an item for the model presented above. This formula is appropriate for use with equal or unequal base rate (Meehl & Rosen, 1955) discrimination problems.

$$D = .5 + \left(\frac{|(.5B + .5) - [(PP_1 + QQ_1) + B(p_2 + qq_2)]|}{(B + 1)} \right)$$

D = proportion correctly identified

$B = N$ for group B/ N for Group A

P = Group A probability of $r(s)_1$

p = Group B probability of $r(s)_1$

$Q = 1 - P$

$q = 1 - p$

P_1 = probability of R_1 for the A group $r(s)_1$ persons

p_2 = probability of R_2 for the B group $r(s)_1$ persons

Q_1 = probability of R_1 for the A group $r(s)_2$ persons

q_2 = probability of R_2 for the B group $r(s)_2$ persons

Conclusions concerning interpretive ambiguity and discriminating power. The generalized item discriminating power formula presented above was used to systematically study the influence of base rate changes and changes in interpretation and response probabilities on item discriminating power. The primary conclusions from this study are discussed below.

1. Interpretive ambiguity in itself does not place a limit on item discriminating power. If the mediating interpretations are each perfectly discriminating, the item will be perfectly discriminating regardless of the level of ambiguity. (If $P_1 = Q_1 = p_2 = q_2 = 1.0$; $D = 1.0$ for all values of P and p .) This means that *freedom from interpretive ambiguity is not a necessary assumption of personality inventories*, and research aimed only at describing the extent of interpretive ambiguity in personality inventories is irrelevant to personality inventory utility.

2. If the mediating interpretations are equal in discriminating power, and the majority response is the same for both interpretations, changes in interpretive ambiguity have no effect on item discriminating power. (If $P_1 = Q_1$ and $p_2 = q_2$, D is constant for all values of P and p .) However, in the special case where neither interpretation is highly discriminating and where the majority response is different for the two interpretations (e.g., $P_1 = 1$; $p_2 = 0$; $Q_1 = 0$; $q_2 = 1$) an increase in item ambiguity which leads to different interpretation probabilities for the two groups will lead to increased discriminating power.

3. When the interpretations differ in discriminating power, but the discrimination direction is the same for both interpretations (the direction of the relations between P_1 and $(1 - p_2)$ and Q_1 and $(1 - q_2)$ are the same), then:

(a) Increasing item clarity by increasing the probability of the most discriminating interpretation (for the group most discriminated by that interpretation) yields a linear increase in item discriminating power. Table 1 interpretation probability values increasing from .5 through 1.0 give examples of these increases in item discriminating power for several levels of interpretation discriminating power.

(b) Increasing item clarity by decreasing the probability of the most discriminating interpretation (for the group most discriminated by that interpretation) yields a linear decrease in discriminating power. Table 1 interpretation probability values decreasing from .5 through .0 provide examples.

These computations suggest an explanation of Elias' data showing ambiguous items as more discriminating than unambiguous items (1951). It is hypothesized that the superiority of Elias' ambiguous items was a function of the increased probability of a more discriminating interpretation rather than a result of the ambiguity itself.

4. The discriminating power of the most differentiating interpretation is the maximum

TABLE 1

CHANGE IN ITEM DISCRIMINATING POWER FOR VARIOUS LEVELS OF INTERPRETATION DISCRIMINATING POWER AND INTERPRETATION PROBABILITY

($p = p_2 = q_2 = .5$; $B = 1$)

Interpretation Probability (P)	Item Discriminating Power ^a		
	$P_1 = .6$	$P_1 = .8$	$P_1 = 1.0$
.0	.50	.50	.50
.2	.51	.53	.55
.4	.52	.56	.60
.5	.525	.575	.625
.6	.53	.5	.65
.8	.54	.62	.70
1.0	.55	.65	.75

^a If $P_1 = .5$, item discriminating power will equal .5 for all values of P .

potential discriminating power of an item. This maximum discrimination level is approached as item wording changes increase the probability of this most discriminating interpretation. This conclusion suggests a change in the usual inventory construction procedure of empirical selection of statements from an a priori item pool. Increasing discriminating power would be possible by empirical selection among the various statement wordings suggested by interpretations reported by respondents. (The exception to this general rule is the special case under 2 above.)

5. When mediating interpretations differ in discrimination direction, changes in the probability of an interpretation can cause a shift in the discrimination direction of the item. For example, consider the MMPI item "I know who is responsible for most of my troubles." This statement has two common interpretations. One is to include one's self as a causal factor. The other is to interpret the item as referring only to others as causing one's troubles. Let us assume, as seems reasonable, that the first interpretation (including one's self) yields a greater percentage of Yes responses in normals than among paranoids. (Assuming $P_1 = .5$, $p_2 = .2$.) Because of the paranoid's greater tendency to project causation, it is likely that the second interpretation (excluding one's self) yields a greater percentage of Yes responses among paranoids. (Assume $Q_1 = .8$, $q_2 = .5$) MMPI respondent reports suggest that as the item is now worded, the interpretation of excluding one's self as a causal factor has the higher probability. In an equal base rate population a Yes respondent should then be classed as abnormal. (This is what is done in practice. A Yes respondent to this item is given an additional point on the Pa scale.) However, if a clinician is disturbed by variance in the interpretations reported by his patients, and "clarifies" the item by changing its wording to "I am responsible for my troubles," because of a belief that "insight" is the variable the test constructor had been trying to get at, the discrimination direction of the item would shift. After this wording change, Yes respondents should be labeled as normal. The relation of discrimination power and direction to interpretive ambiguity for the probability

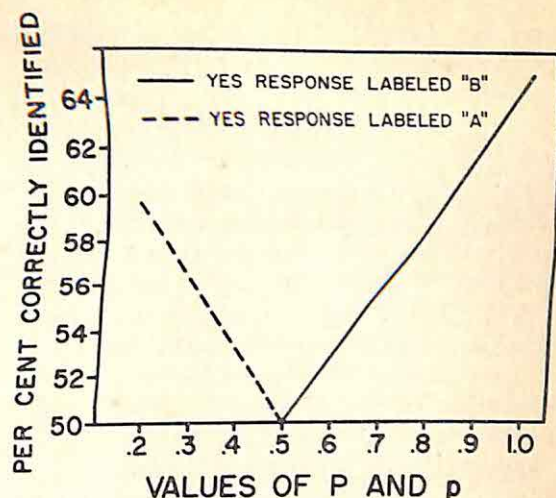


FIG. 1. Shift in discrimination direction occurring as a function of ambiguity lessening changes in item wording. ($B = 1$, $P_1 = .5$, $Q_1 = .8$, $p_2 = .2$, $q_2 = .5$; P and p values of 1.0 indicate an unambiguous item.)

values used in this example is illustrated in Fig. 1.

School and other institutional administrators occasionally suggest item wording changes before granting approval to use subjects for research projects. Although the changes suggested usually seem minor, the shifts in discrimination direction, which can be caused even by "clarifying" wording changes, make such changes a very dubious research procedure.

Taylor's MAS wording changes (1953) are another example of clarifying items without adequate study of resultant shifts in item discriminating power and direction. Taylor tested the comparability of the old and revised MAS forms by studying their correlation in a college population. Taylor states the rationale for the use of this population as follows:

A sample was selected from the college population for this purpose since it was thought that this group would show the least confusion in interpreting the original versions of the difficult items and, therefore, better demonstrate the comparability of the two forms than less verbally sophisticated individuals (Taylor, 1953, p. 288).

Wording changes designed to elicit a greater probability of the high intelligence group interpretation of the original items (judges selected these revised items for nearness to their

interpretation of the meaning of the original item) were thus tested for comparability to the older items on a group explicitly selected for a high probability of the same interpretation of both item forms. Such a method certainly does not provide an adequate test of the comparability of original and revised test forms if the test is to be used in more general population samples. This item revision and revision testing method has not influenced research conclusions based on the MAS because the great bulk of research has been done using only the revised form. However, the same type of revision and revision testing done on a personality inventory with an interpretive base grounded on research with the initial form, would be likely to lead to incorrect conclusions if the old form interpretations were carried over to the revised form. These comments apply to demonstrations of the comparability between any two test forms. *Whenever there is a possibility of interpretive ambiguity, and interpretation choices are not assumed to be random, the comparability of two forms of the same test must be demonstrated on a population with relevant base rates comparable to the base rates in the population in which the tests are to be used.*

6. Use of personality inventories in populations with different base rates can change item discrimination direction and power. For example, at P and p levels of .5 for the item used in Fig. 1, 55% correct identification can be obtained with a $\frac{2}{3}$ A group base rate by labeling Yes respondents as A. However, with a $\frac{2}{3}$ B group base rate, Yes respondents should be labeled as B, and with a .5 base rate for each group, the item is nondiscriminating. These considerations are especially important because of the almost invariable base rate changes between test development populations and populations in which the test is used (Broen, 1957). Also, *unless the interpretation probabilities (P , p) and interpretation response probabilities (P_1 , p_2 , Q_1 , q_2) are known for each group, the effects of known base rate changes on item discrimination direction and power cannot be computed.* Three alternatives are then open to test constructors and validators: (a) accept introspective reports of item interpretations as valid self-reports and compute interpretation and in-

terpretation response probabilities from these introspective reports; (b) study item discriminating power in populations with relevant base rates comparable to those in the population in which the test is to be used; (c) use unambiguous (single interpretation) items. The interpretation response probabilities for the group to be discriminated can, of course, be easily computed for unambiguous items. The utility of using respondent reports as ambiguity measures (in c) and interpretation discriminators (in a) need not be only an assumption. The utility of the respondent reports can be demonstrated by research dealing with the accuracy of discriminating power predictions using the probability values obtained through the respondent reports for populations with various base rates. Such research would also yield data relevant to Vernon's (1953, p. 140) assumption that the interpretations are distributed randomly throughout various diagnostic classifications. The greater ease of (c) over (a) and the greater applicability to all populations with known base rates of (c) over (b) may be the strongest argument for freedom from ambiguity as a criterion in item selection.

Response Ambiguity

The second type of item ambiguity—response ambiguity—has been defined above as the extent to which groups to be discriminated give similar responses and is, therefore, the inverse of item discriminating power. Because of this direct relation to item validity, personality inventory criticisms based on response ambiguity (Allport, 1937; Anastasi, 1954; Eisenberg, 1941; Freeman, 1955; Rotter, 1954) would seem to be more directly relevant to personality inventory utility than criticisms based on interpretive ambiguity. However, methods for combining items of relatively low discriminating power to make highly discriminating scales are available. These methods involve using item configurations (Meehl, 1950); using nondiscriminating items as suppressors of irrelevant variance (Horst, 1941); or combining items which have somewhat higher correlations with the criterion than with each other. An example of the latter type of scale is provided by the

MMPI *Hy* scale which combines two relatively uncorrelated sets of items, each of which provides increased discrimination to the total scale by differentiating on separate dimensions (McKinley & Hathaway, 1944). These methods which enable the use of items of low criterion discriminating power to construct highly discriminating scales indicate that, like freedom from interpretive ambiguity, freedom from response ambiguity is not a necessary assumption for personality inventories.

SUMMARY

An ambiguity probability model was introduced to clarify the relation of personality inventory interpretive ambiguity to discriminating power. Item discriminating power was found to be a function of interpretation probability, interpretation discriminating power, and population base rates. Research and test construction implications of these findings were discussed, and several methods of using relatively nondiscriminating test items to increase scale discriminating power were cited.

REFERENCES

- ALLPORT, G. W. *Personality: A psychological interpretation*. New York: Holt, 1937.
- ANASTASI, ANNE. *Psychological testing*. New York: Macmillan, 1954.
- BENTON, A. L. The interpretation of questionnaire items in a personality schedule. *Arch. Psychol.*, N. Y., 1935, No. 190.

- BROEN, W. E., JR. The use of psychometric sign tests in different populations. *Psychol. Rec.*, 1957, 7, 117-118.
- EISENBERG, P. Individual interpretation of psychoneurotic inventory items. *J. gen. Psychol.*, 1941, 25, 19-40.
- ELIAS, G. Self-evaluation questionnaires as projective measures of personality. *J. consult. Psychol.*, 1951, 15, 496-500.
- FREEMAN, F. S. *Theory and practice of psychological testing*. New York: Holt, 1955.
- HORST, P. The prediction of personal adjustment. *Soc. Sci. Res. Coun. Bull.*, 1941, No. 48, 1-156.
- HUTT, M. L. The use of projective methods of personality measurement in army medical installations. *J. clin. Psychol.*, 1945, 1, 134-140.
- LANDIS, C., & KATZ, S. E. The validity of certain questions which purport to measure neurotic tendencies. *J. appl. Psychol.*, 1934, 18, 343-356.
- McKINLEY, J. C., & HATHAWAY, S. R. The MMPI: V. Hysteria, hypomania and psychopathic deviate. *J. appl. Psychol.*, 1944, 28, 153-174.
- MEEHL, P. E. The dynamics of "structured" personality tests. *J. clin. Psychol.*, 1945, 1, 296-303.
- MEEHL, P. E. Configural scoring. *J. consult. Psychol.*, 1950, 14, 165-171.
- MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. *Psychol. Bull.*, 1955, 52, 194-216.
- ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- VERNON, P. E. *Personality tests and assessments*. New York: Holt, 1953.
- WELSH, G. S., & DAHLSTROM, W. G. *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.

(Received May 4, 1959)

IMPLICIT PERSONALITY THEORIES OF CLINICIANS AS DEFINED BY SEMANTIC STRUCTURES¹

MAURICE KORMAN

Southwestern Medical School, University of Texas

The consistently poor showing of clinicians in large scale global assessment studies (Kelly & Fiske, 1951; Holtzman & Sells, 1954) has led to increased concern over the bias introduced by constant processes in the assessor, his reliance upon the connotative meaning of key terms and his idiosyncratic use of concepts.

In his analysis of certain judgmental variables, Cronbach (1955) has shown that if it is a judge's expectation that certain traits covary and others not, he will customarily report subjects to be high (or low) on all of a set of traits. Such a tendency, suggested to reflect an assessor's implicit personality theories, has been demonstrated in different types of judges in diverse settings (Jones, 1954; Soskin, 1954; Steiner, 1954). The psychologist, psychiatrist, or social worker who first observes, tests, or interviews and then applies labels or makes decisions is especially deserving of such attention to his semantic space since many of the concepts he uses are characterized by the looseness and ambiguity of their definitions (Grayson & Tolman, 1950) and their lack of integration in a testable nomological network (Feigl, 1949). Moreover, since the multidisciplinary "team approach" to the patient has found wide acceptance, the lack of semantic clarity and precision has been aggravated by problems of interdisciplinary communication. By virtue of

their different training and interests, and because of the partly distinct role they assume, it would seem reasonable to expect psychiatrists, social workers, and psychologists to base their inferences on distinct types of implicit personality theories.

In this study, the clinician's semantic structure is used to reflect that system of interrelationship of meanings that forms one important aspect of his implicit personality theories. This model, based on Osgood and Suci's *D* measure (1952), represents similarities in connotative meaning between concepts as discriminated by an appropriate form of the semantic differential. As applied to groups, this technique has excellent reliability (Jenkins, Russell, & Suci, 1957) and the model which is based upon it can be described as having good construct validity (Lazowick, 1955; Mowrer, 1953; Osgood, 1952).

This investigation will attempt to deal separately with interstructure differences (between disciplines) and intrastructure differences (among concepts). Two questions will be asked: 1. How similar are the implicit personality theories of psychiatrists, psychologists, and social workers, and with respect to what dimensions do they differ? 2. Does an intensive examination of such a network of associations suggest the existence of meaningful covariations or clusterings of concepts?

METHOD

Subjects

The subjects (Ss) were 10 psychiatrists, 10 psychiatric social workers, 10 clinical psychologist trainees, 10 PhD psychologists. The first three groups were chosen from among advanced graduate students or psychiatric residents who were near achieving that arbitrary level of training and experience which their

¹ This paper, presented at the meeting of the Southwestern Psychological Association in Topeka, 1959, represents a portion of a dissertation submitted to the University of Minnesota in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is greatly indebted to his advisor, William Schofield, and to others whose assistance aided in the completion of this study.

TABLE 1
CONCEPTS AND SCALES USED IN THIS ANALYSIS

Concepts (listed in the order of their frequency)		Scales (listed by factors)		
Diagnostic	Therapeutic	Evaluation	Potency	Activity
Anxiety Reaction	EST (Electroshock Therapy)	Good-Bad	Hard-Soft	Active-Passive
Depressive Reaction	Psychotherapy	Kind-Cruel	Masculine-Feminine	Excitable-Calm
Schizoid Personality	Supportive Psychotherapy	Successful-Unsuccessful	Strong-Weak	Fast-Slow
Schizophrenic Reaction	Tranquilizing Drugs			
Schizophrenic Reaction, Paranoid Type	Reassurance			
Somatization Reaction	Uncovering Psychotherapy			
Hysteria	Commitment			
Psychopathic Personality	Group Psychotherapy			
Paranoid Personality	Directive Psychotherapy			
Obsessive-Compulsive Reaction				
Passive-Dependent Personality				

respective disciplines demand of them before they are accorded full professional status. The fourth group, composed of senior clinical psychologists who held the PhD degree for an average of 5.80 years, was added in order to further test the sensitivity of the techniques employed and to increase the number of psychologists to a more stable *N* for the investigation of that group's semantic structure.

Concepts

The terms used in this study represent a sample from an empirically derived pool of 95 terms based on an examination of 80 medical folders from four Minneapolis Psychiatric In- and Out-Patient Services. Of the 31 terms or phrases relating to therapy which were thus recorded, 9 were selected from among the 13 most frequently used ones by discarding 4 somewhat ambiguous (e.g., Active Psychotherapy) or less widely used concepts (e.g., Regression EST). The 64 diagnostic terms in this pool were likewise ranked in the order of their total frequency of occurrence. The 10 most frequently used diagnostic terms were included in this study, together with Passive-Dependent Personality which was used in spite of its lower frequency count because of its contrast to the other terms. The concepts and scales used in this investigation are reported in Table 1.

Scales

Of the many scales available from the various factorial studies, none have been found to be representative of one factor only, and it has been suggested that the factor loadings of the scales vary with the concepts being rated (Osgood, 1952). Since the *Ss* were busy professionals, it was decided to restrict the number of scales to nine to be chosen from the pool of 20 scales that had been set up for the Minnesota *Atlas* study (Jenkins, Russell, & Suci, 1957) and which are believed to be representative of the major dimensions of meaning. A pilot study determined which of these scales best differentiated this type of concept from one another. The 9 scales chosen for this study were, with one exception, the scales having the largest variance. Although Colorful-Colorless was a more widely used scale than Masculine-Feminine, it was decided to include the latter instead in order to give the three major factors, evaluation, potency, and activity, equal representation in terms of number of scales per factor.

Procedure

After the order of the concepts and of the scales had been randomized, and the poles of five of the nine scales reversed so that scales with the same factor loadings were systematically counterbalanced, the

40 Ss were instructed to rate each concept on the nine scales using the entire continuum in reporting their first impression of each term.

RESULTS

Intergroup Differences

Each group's semantic structure was determined as follows. Each group's mean scale profile for each concept was determined and as a measure of the similarity in connotation of meaning of any pair of concepts, Osgood and Suci's *D* was obtained. *D* is defined as the square root of the sum of the square differences between scores on each of the scales for any two concepts. Each group's semantic structure is represented by the matrix of *D*s reflecting the distances between the concepts used. Since the use of Pearson product-moment correlations to gauge comparative similarity in semantic structures has been shown to be a sensitive measure by Osgood and Luria (1954) the correlations between each group's set of *D* values and that of every other group were obtained and are reported in Table 2. As can be seen, none of the intercorrelations between group semantic structures falls below .60. All correlations are significantly different from zero ($p < .001$).

In testing the significance of the differences between product moment correlations reported in Table 2, it was found that the semantic structures of the psychology trainees and the PhD psychologists are significantly closer than (a) the psychology trainees and the social workers ($p < .01$), (b) the psychology trainees and the psychiatrists ($p < .03$) and (c) the PhD psychologists and the social workers ($p < .01$). Moreover, the semantic

TABLE 2
INTERCORRELATIONS OF GROUP *D* MATRICES

	PhD Psychol- ogists	Psychology Trainees	Psychi- atrists
Psychology trainees	.78		
Psychiatrists	.74	.69	
Social workers	.61	.60	.73

TABLE 3
NUMBER OF *D*s DIFFERENTIATING THE FOUR GROUPS
AT THE .05 LEVEL OF SIGNIFICANCE¹

	PhD Psychol- ogists	Psychol- ogy Trainees	Psychi- atrists
Psychology trainees	1(1,0)		
Psychiatrists	1(0,1)	1(0,1)	
Social workers	7(5,2)	10(5,5)	3(2,1)

¹ The number of *D*s involving diagnostic concepts are given first within the parentheses, followed by the number of therapeutic *D*s.

structures of the PhD psychologists and of the psychiatrists are significantly closer to one another than the semantic structures of the PhD psychologists and the social workers ($p < .02$).

In an effort to explicate the semantic differences between social workers and the other disciplines, the four groups were contrasted by means of Mann Whitney *U* Tests on each of the 91 *D*s or "links" constituting their semantic structure. As can be seen from Table 3, 23 *D*s significantly differentiated the four groups from one another at the .05 level of significance. Twenty of these differences are between social workers and the other groups; in contrast, the psychologists and psychiatrists differ significantly on only 3 *D*s. Twelve of the 20 *D*s (60%) involve diagnostic categories while the remaining 8 concern therapeutic terms. Since exactly 60% of the pool of 91 *D*s relate to diagnostic terms, it can be seen that social workers differ from the other disciplines as much with regard to diagnostic concepts as they do with reference to therapeutic phrases. A *U* Test applied to the data of Table 3 indicates that the number of significant differences between social workers and the other disciplines is itself significantly higher than the number of such differences among the remaining three groups ($p = .05$). Table 4 reports specific differences between social workers on the one hand and psychologists and psychiatrists on the other. It should be noted that except for Obsessive-Compulsive/Depressive Reaction the social workers find these pairs of con-

TABLE 4

PAIRS OF CONCEPTS DIFFERENTIATING THE GROUPS AT THE .05 LEVEL OF SIGNIFICANCE

Social Workers vs. Other Three Groups ^a		PhD Psychologists vs. Psychology Trainees vs. Psychiatrists	
Diagnostic	Therapeutic	Diagnostic	Therapeutic
Hysteria/Obsessive-Compulsive Reaction	Directive PRx ^b / Tranquilizing Drugs	Hysteria/ Somatization reaction	Reassurance/ Supportive PRx
Passive-Dependent Personality/Paranoid Personality	Group PRx/ Tranquilizing Drugs	Directive PRx/EST	
Obsessive-Compulsive reaction/ Psychopathic Personality	Uncov. PRx/ Tranquilizing Drugs		
Somatization Reaction/ Paranoid Personality	EST/Tranquilizing Drugs		
Hysteria/Somatization Reaction	EST/Supportive PRx		
Hysteria/Psychopathic Personality			
Somatization reaction/ Psychopathic Personality			
Obsessive-Compulsive reaction/ Depressive Reaction			

^a Only 13 out of 20 significant differences are reported here since 5 of the *Ds* listed differentiated the social workers from more than one other group.

^b Psychotherapy.

cepts to be *closer* in connotative meaning than do their colleagues.

The Semantic Structure of Psychologists

Since relatively minor differences were expected between the PhD psychologists and the psychology trainees, these groups were combined in order to add more stability to the results. Because the statistical study of relational measures such as *D* is cumbersome and requires many separate tests of significance for even a small number of concepts, only the nine therapeutic terms were used in this phase of the investigation.

The 36 *Ds* between all pairs of the nine therapeutic terms for each of the 20 psychologists were computed, resulting in 20 matrices or 36 sets of *Ds* each with an *N* of 20. Each concept can be visualized as a point a certain distance away from each of the other terms and the significance of the difference in distance between different pairs of concepts can be calculated by using each concept in turn as a point of reference. In order to answer the kind of question, "Is

supportive therapy closer to reassurance than it is to directive psychotherapy," it was necessary to compute 252 Mann Whitney *U* Tests in all. Of these, 83 were significant at the .05 level, 65 being significant at the .02 level.

The presence of such a large number of significant differences in the distance between therapeutic concepts suggests the existence of clusters of concepts. Examination of the sta-

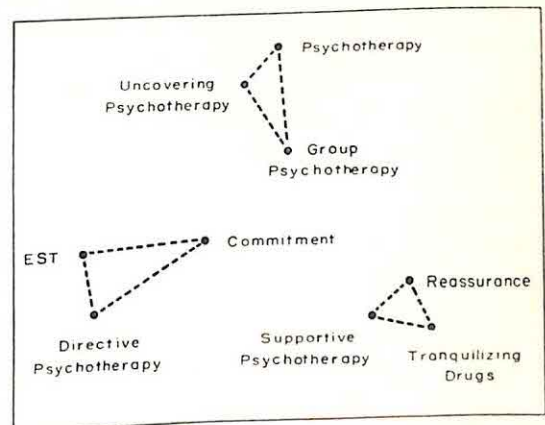


FIG. 1. A two dimensional representation of clusters of therapeutic terms.

tistically significant *Ds* between the nine therapeutic terms indicates the existence of three fairly distinct clusters: (a) EST, Directive Psychotherapy, and Commitment; (b) Uncovering Psychotherapy, Psychotherapy, and Group Psychotherapy; (c) Reassurance, Supportive Psychotherapy, and Tranquilizing Drugs. Figure 1 is a two-dimensional visualization of the interrelationships based primarily on topological relations as defined by the tests of significance between *Ds*.

DISCUSSION

It is apparent from Tables 2, 3, and 4 that the greatest intergroup differences in implicit personality theories lie between psychologists (regardless of level of experience) and social workers, with psychiatrists maintaining a mediating position between these two disciplines. It must, however, be emphasized that the design of this study did not allow for the testing of the overall importance of these interdisciplinary differences; compared, for instance, to the differences that might possibly emerge from the contrasting of these groups with a group of laymen, such differences as were found in this study might seem minimal and insignificant.

The position of the psychology trainees with respect to the senior psychologist is as expected. There is a communality of interests and outlook between these groups made all the more pervasive by the fact that in this particular sample one group trains and supervises the other. The relationship of both groups of psychologists to the social workers is likewise as hypothesized. Of the three major interests of psychologists, psychodiagnostics, research, and psychotherapy, only the last is shared in some measure by social workers, whose major concern has traditionally been the social welfare of the patient. That psychiatrists should be as "close" to the senior psychologists as to the social workers had not been predicted. One possible explanation would emphasize the fact that the psychiatrist shares both the interest in diagnosis and therapy (as well as in research to an increasing extent) of the psychologist, and the social worker's concern with the patient's family, job, financial status, etc. One impor-

tant inference can further be drawn from these results: Since gross intergroup similarities and differences predicted on the basis of these groups' interests and training are generally reflected in interdisciplinary differences in semantic structures and implicit personality theories, some validity is given to the conclusion that semantic differential ratings and the model upon which the semantic structure is based are sensitive tools and that they parallel extralinguistic interests and attitudes.

In examining the pairs of concepts which social workers find more connotatively similar (Table 4), two alternative explanations offer themselves: (a) the two words are actually closer in connotative meaning for social workers or (b) social workers do not *discriminate* between the terms, as when all diagnostic terms are uniformly rated "somewhat bad." The second alternative seems the likeliest inasmuch as the pairs of concepts which are connotatively closer for social workers seem to involve constructs that are generally thought to be quite distinct in meaning (e.g., Obsessive-Compulsive reaction/Psychopathic personality). Since these are the only three such differences between PhD psychologists, psychology trainees and psychiatrists, it is difficult to decide whether they represent opposite and interpretable attitudes or chance differences. If these are real differences, the content of the three pairs of concepts (e.g., Hysteria/Somatization reaction) would suggest arguable theoretical differences rather than lack of discrimination. In either case, the semantic structures of psychologists and psychiatrists are seen not to be appreciably dissimilar. The conclusion may, therefore, be drawn that insofar as these semantic structures represent an aspect of the Ss' implicit personality theories, the systems of covariation of diagnostic and therapeutic terms used by psychiatrists and psychologists are much the same. Social workers, on the other hand, reveal attitudes and associations not shared by psychologists, although accepted to a larger extent by psychiatrists.

Although the intensive analysis performed on the interrelationships among nine terms relating to therapy was based on the semantic structures of psychologists only, it can be seen from the above considerations that the results

could be generalized to the psychiatrists. The first question that presents itself in this attempt to tap some aspects of the clinician's implicit personality theories concerns the basis for the particular clustering that was found. What dimensions account for the intracluster similarities? If one considers the overall median factor scores² of terms relating to therapy separately for each of the three factors represented in this study (evaluation, potency, and activity) it becomes clear that Cluster 3 (Reassurance, Supportive Psychotherapy, Tranquilizing Drugs) is "valued" more highly than Cluster 1 (Directive Psychotherapy, Commitment, EST); Cluster 1, however, is more potent or stronger as well as more active. Cluster 2 (Psychotherapy, Uncovering Psychotherapy, Group Psychotherapy) is not as cohesive as the other clusters and it tends to share in the characteristics of the other two groupings.

What attitudes and implicit beliefs relevant to the therapies most commonly prescribed for psychiatric patients can be inferred from this data? It might be argued that the terms in each cluster represent treatments of choice under certain specifiable conditions, other things being equal. Cluster 1, the "strong, active" therapies, might be most applicable to most situations which to the mind of the clinician demanded such rapidly direct and active therapeutic interference at the cost of being forced to sacrifice more highly regarded or valued therapies. Cluster 3, the "good, kind" therapies, would be chosen when a nearly opposite situation existed: the emphasis required would be on a "good, kind, successful" therapy involving relatively little direct interference with the patient's psychopathology. The third cluster, because of its middle position on the three factors, could be said to represent a compromise between more valued therapies and more active and strong ones. Where neither support and reassurance nor active interference are required, Psychotherapy, Uncovering Psychotherapy, and Group Psychotherapy would then seem to be indicated. These terms also differ relatively more among themselves than is the

case with the other two clusters. Clinicians seem to attach a valuatative connotation to *Psychotherapy* that they find lacking in the other two terms. Is it possible that they have in mind a type of psychotherapy (the one they practice, perchance?) which they prefer to the ones listed here?

The value of such inferences must be tested primarily from their agreement with other knowledge. Insofar as these interpretations clarify the decision making activities of clinicians, and insofar as clinical practice in turn lends credence to them, they contribute to our knowledge of clinicians. Unfortunately, no systematic data is available that reports on clinicians' attitudes towards these therapies. One is therefore forced to fall back on such weak criterion as "do these conclusions make sense in light of how clinicians are commonly believed to feel about these therapies?" The present writer would suggest that extralinguistic attitudes and beliefs concerning therapy are indeed being tapped which are in accordance with the way in which clinicians implicitly match a therapy with their "feeling about the patient." Nevertheless, it cannot be too strongly emphasized that this is an exploratory study in an area in great need of more precise and systematic investigation.

SUMMARY AND CONCLUSIONS

In an effort to investigate the implicit personality theories of clinicians, 10 clinical psychology trainees, 10 psychiatrists, 10 social workers, and 10 senior clinical psychologists were asked to rate 11 diagnostic and 9 therapeutic frequently used concepts on nine appropriate semantic differential scales. The resulting semantic structures define an aspect of the groups' implicit personality theories. Major conclusions were as follows:

1. Significant intergroup differences in semantic structures and in degree of connotative similarity assigned to pairs of concepts exist between psychologists (regardless of level of experience) and social workers, with psychiatrists maintaining a mediating position between these two disciplines. It is suggested that social workers make fewer discriminations between the connotative meaning of psychiatric and psychological terms.

² A thesaurus of diagnostic and therapeutic terms based on factor scores is available from the author.

2. An intensive analysis of the interrelationships between nine terms relating to therapy for 20 psychologists suggests the existence of three clusters of terms: (a) Directive Psychotherapy, Commitment and EST, (b) Psychotherapy, Uncovering Psychotherapy and Group Psychotherapy, and (c) Reassurance, Supportive Psychotherapy, and Tranquilizing Drugs. These clusters (a) can be said to represent the degree of equivalence and interchangeability of the terms in the cluster and (b) are suggestive of situations in clinical practice in which the terms of one cluster would be more applicable than those of another. Such clusterings can be presented as providing an insight into the implicit personality theories of clinicians.

REFERENCES

- CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, 52, 177-193.
- FEIGL, H. Operationism and scientific method. In H. Feigl & W. Sellers (Eds.), *Readings in philosophical analysis*. New York: Appleton-Century-Crofts, 1949.
- GRAYSON, H. M., & TOLMAN, RUTH S. A semantic study of concepts of clinical psychologists and psychiatrists. *J. abnorm. soc. Psychol.*, 1950, 45, 216-231.
- HOLTZMAN, W. H., & SELLS, S. B. Prediction of flying success by clinical analysis of test protocols. *J. abnorm. soc. Psychol.*, 1954, 49, 485-490.
- JENKINS, J. J., RUSSELL, W. A., & SUCI, G. J. An atlas of semantic profiles for 360 words. *Univ. of Minnesota Studies on the Role of Language and Behavior*, 1957, Tech. Rep. No. 15.
- JONES, E. E. Authoritarianism as a determinant of fast impression formation. *J. Pers.*, 1954, 23, 107-127.
- KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univ. Michigan Press, 1951.
- LAZOWICK, L. On the nature of identification. *J. abnorm. soc. Psychol.*, 1955, 51, 175-183.
- MOWRER, O. H. Changes in verbal behavior during psychotherapy. In O. H. Mowrer (Ed.), *Psychotherapy, theory and research*. New York: Ronald, 1953.
- OSGOOD, C. E. Report on development and application of the semantic differential. Unpublished manuscript, Univ. of Illinois, 1952. (Mimeo.)
- OSGOOD, C. E., & LURIA, Z. A blind analysis of a case of multiple personality using the semantic differential. *J. abnorm. soc. Psychol.*, 1954, 49, 579-591.
- OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean differences and profile information. *Psychol. Bull.*, 1952, 49, 251-262.
- SOSKIN, W. F. Frames of reference in personality assessment. *J. clin. Psychol.*, 1954, 10, 107-113.
- STEINER, I. D. Ethnocentrism and tolerance of trait "inconsistency." *J. abnorm. soc. Psychol.*, 1954, 49, 349-354.

(Received May 19, 1959)

WISC IQs FOR THE MENTALLY RETARDED

DONALD P. OGDON

Richmond Professional Institute, College of William and Mary

Many psychologists using the Wechsler Intelligence Scale for Children (WISC) have been chagrined to discover that a client has earned an insufficient number of weighted score units to obtain an IQ from the tables provided: the lowest full scale IQ given in the WISC manual is 46, which requires a scaled score of 26 (Wechsler, 1949, p. 26). The primary purpose of this paper is to provide extrapolated IQs down to 28, which is obtained when the scaled score is equal to one.

This extrapolation provides IQs to within the Idiot range, according to Wechsler's (1958, p. 57) classification system. For a variety of reasons, it is considered important by psychologists and others working with the mentally retarded to differentiate among them. Many institutional policies and state legislation require a diagnosis of level of mental retardation or deficiency. This obviously cannot always be provided with the published WISC data. IQs often play a significant role in determining differential therapy, treatment, and training or education in these institutions. However, the principal and most valid reason for desiring to differentiate among in-

dividuals with IQs below 46 is to further research and understanding of them. The potential positive contributions to society of these individuals are only beginning to be recognized (Hutt & Gibby, 1958, p. 99).

In Fig. 1, the Full Scale WISC IQs have been plotted against WISC scaled scores. This resulted in a straight line which can be described by the formula $IQ = (\text{scaled score}) (.727) + 27.4$. The formula gives results, when rounded to the nearest whole IQ value, which are identical with those published in the WISC manual for every IQ between 46 and 85 with

TABLE 1
EXTRAPOLATED WISC IQs: $IQ = (\text{SCALED SCORE})$
 $(.727) + 27.4$

Scaled Scores	Calculated IQ Values	Smoothed IQ Values
25	45.58	46
24	44.85	45
23	44.12	44
22	43.39	43
21	42.67	43
20	41.94	42
19	41.21	41
18	40.49	40
17	39.76	40
16	39.03	39
15	38.31	38
14	37.58	38
13	36.85	37
12	36.12	36
11	35.40	35
10	34.67	35
9	33.94	34
8	33.22	33
7	32.49	32
6	31.76	32
5	31.04	31
4	30.31	30
3	29.58	30
2	28.85	29
1	28.13	28

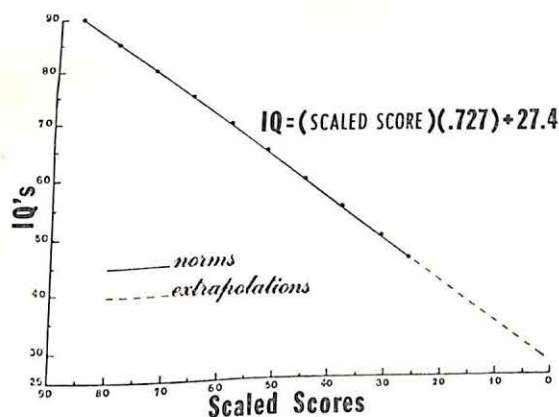


FIG. 1. WISC full scale IQs and scaled scores with extrapolated points below norms.

only one exception—with a scaled score of 69, the formula gives the calculated IQ value of 77.56, while Wechsler's Table XI B (1949, p. 26) gives an IQ of 77.

Table 1 presents extrapolated WISC IQs for scaled scores from 1 to 25.

These extrapolated IQ values may be used appropriately only with certain cautions or limitations. Firstly, although the empirical relationship is known to be linear, some relationships that are linear "in the middle range" of their distributions depart from linearity at the extremes.

Generally, the further the extrapolation is made from the empirical data, the greater the probability of error. In this case, as our extrapolated IQs decrease from 46, the less con-

fidence we have in our extrapolation, the lower the expected reliability of the value, and the less its probable psychological significance.

Also attenuating the reliability of the lower extrapolated IQs is the smaller number of items successfully completed to establish the scaled score.

REFERENCES

- HUTT, M. L., & GIBBY, R. G. *The mentally retarded child*. Boston: Allyn-Bacon, 1958.
- WECHSLER, D. *The measurement and appraisal of adult intelligence*. (4th ed.) Baltimore: Williams & Wilkins, 1958.
- WECHSLER, D. *Wechsler Intelligence Scale for Children*. New York: Psychological Corp., 1949.

(Received March 31, 1959)

INCIDENCE OF BENDER-GESTALT FIGURE ROTATIONS

RICHARD M. GRIFFITH AND VIVIAN H. TAYLOR

Veterans Administration Hospital, Lexington, Kentucky

One striking distortion in the drawing of the Bender Visual Motor Gestalt Test (BG) or in the reproduction of block designs—the rotation or angular displacement of one or more of the figures—would be more meaningful to the diagnostician if he had statistics on the frequency of the phenomena in his clinical groups and if he had an adequate theory to elucidate its significance. Aimed at the second problem, a series of attempts to furnish a theoretical underpinning through investigating factors which produce rotations was begun by Shapiro (1953) followed by others (Williams, Lubin, Gieseeking, & Rubenstein, 1956; Yates, 1956) who used similar techniques and, most recently, by Hannah (1958) who worked independently. Earlier, Hanvik and Anderson (1950) had considered the effect of focal brain lesions on the production of rotations. The present note provides actuarial statistics on the first problem—the frequency of rotations in the BG drawings of different clinical groups.

METHOD

Drawer by drawer, the files of folders accumulated through eleven years of testing by the Psychology Service of a neuropsychiatric hospital were searched until approximately 1000 BG records (probably one third of all those available!) were examined. The average age of a restricted sample (331 cases) was 35.5 (*SD*, 11.2) at the time the test was administered. The distribution of ages was skewed by the WWI veterans, the median age being 33.4. The IQs of 282 of these patients averaged 94.9 (*SD*, 16.4). As standard operating procedure, looking ahead to future research possibilities, the diagnosis which had been reached by a team composed of psychiatrist, psychologist, and social worker had been noted in each folder. For the present purposes this diagnosis was accepted without further evidence. It is recognized that the design of the study is to some extent

impure in that the BG, however small its influence, had been one of the considerations in arriving at the diagnosis.¹

The diagnoses were broadly categorized into Schizophrenia, Neurosis, Character Disorder, Chronic Brain Syndrome, Mental Deficiency, and Other, the latter being a catch-pile for the less frequent diagnoses and for those instances for which the diagnosis had inadvertently been omitted. When it was noted that the percentage of rotations was high for the first few mental defectives the number of their records was supplemented by combing other drawers of the files; thereafter, their figures were not used in computing the totals for the hospital population. A rotation was defined as an angular displacement of 45 degrees or more in a figure which otherwise remained recognizable.

TABLE 1
PERCENTAGE OF BG RECORDS WITH
ONE OR MORE ROTATIONS

	<i>N</i>	Percentage with Rotation(s)
Schizophrenia	346	19.7
Neurosis	167	19.8
Character Disorder	165	16.4
Chronic Brain Syndrome	147	40.8*
Mental Deficiency	59	55.9*
Other	178	23.0
Total ^a	1003	22.8

^a Total does not include mental defectives.

* Chronic Brain Syndrome and Mental Deficiency differ at .05 probability level; each differs from all other groups at .01 level.

¹ With some clinicians during much of the period covered by the records the BG had been administered primarily as a buffer to what was considered more valid tests. That its effect on diagnoses was never great may be concluded from the breadth of evidence considered by the diagnostic team and from the fact that before they were told the results of the present study a group of 14 psychodiagnosticians within the hospital estimated almost as high per cent rotations by the schizophrenics (36) as by the organics (43).

RESULTS

Table 1 gives the percentage of the records of each diagnostic group in which at least one of the BG designs had been rotated. To the surprise of the investigators and their colleagues, the percentage for the schizophrenics did not differ materially from those for the neurotics and the character disorders (19.7, 19.8, and 16.4, respectively). Above this cluster of values the rotations by the Chronic Brain Syndrome and the Mental Deficiency groups rose to 40.8 and 55.9%. The 23.0% for Others corresponds closely to the 22.8 for the total population (exclusive of the mentally defective).

DISCUSSION

Hanvik and Andersen (1950), requiring less deviation (30 degrees vs. 45) and employing more stringent diagnostic standards, reported a somewhat larger percentage of rotations by organics. Of a group of 44 patients with focal brain lesions which had been verified either through surgery or by Xray the records of 59.1% had one or more rotations as compared with 40.8% for the present group, which was more heterogeneous. Their 18.9% for 37 normals and neurotics is not out of line with the results for the present NP patients—other than organic and deficient.

Because of the relatively small number of mentally defective patients, their percentage rotations is not considered to be as reliable as those for the other diagnoses. Also, in a

hospital population the diagnosis is always suspect; mental illness may mimic deficiency. That a relation between intelligence and rotations, as conjectured in the literature (Williams et al., 1956; Yates, 1956), does exist, is supported by comparing the IQs of 56 rotators and 226 nonrotators: the means of 84.3 and 97.5, with standard deviations of 13.3 and 17.1, respectively, differ by the *t* test at the .01 level.

SUMMARY

This study, confirming the belief of clinicians that rotations of Bender's gestalten are of diagnostic significance, establishes the relative frequency of such distortions by clinical groups within one neuropsychiatric hospital.

REFERENCES

- HANNAH, L. D. Causative factors in the production of rotations on the Bender-Gestalt designs. *J. consult. Psychol.*, 1958, 22, 398-399.
- HANVIK, L. J., & ANDERSEN, A. L. The effect of focal brain lesions on recall and on the production of rotations in the Bender-Gestalt test. *J. consult. Psychol.*, 1950, 14, 197-198.
- SHAPIRO, M. B. Experimental studies of a perceptual anomaly: III. The testing of an explanatory theory. *J. ment. Sci.*, 1953, 99, 394-409.
- WILLIAMS, H. L., LUBIN, A., GIESEKING, C., & RUBINSTEIN, I. The relation of brain injury and visual perception to block design rotation. *J. consult. Psychol.*, 1956, 20, 275-280.
- YATES, A. J. The rotation of drawings by brain-damaged patients. *J. abnorm. soc. Psychol.*, 1956, 53, 178-181.

(Received May 15, 1959)

BRIEF REPORTS

A FACTOR ANALYTIC STUDY OF THE SELF-CONCEPT¹

PHILIP A. SMITH²

System Development Corporation, Santa Monica, California

Most measures of the self-concept generate scores which are treated as positive or negative points on a value continuum, despite evidence (Osgood, Suci, & Tannenbaum, 1957) which suggests that a meaningful concept such as the "self" comprises an aggregate of factors rather than a single evaluative dimension. The present study, which was undertaken in an attempt to categorize self-ratings in such a way as to permit a sensitive evaluation of change in psychiatric patients during therapy, casts further light on the factorial structure of the self-concept.

A self-rating device composed of 70 bipolar adjectives descriptive of human personality, including antonyms used by Osgood and terms selected from several popular self-concept measures, was completed by a heterogeneous group of 120 young adult male psychiatric patients. Ratings by each S on all 70 scales were dichotomized at the median and a 70 × 70 matrix of phi coefficients was generated. Two overlapping portions of this matrix were separately factor analyzed and rotated by the quartimax method, an analytic procedure (conveniently programmed for the computer) which provides an objective approximation of orthogonal simple structure. Five interpretable factors were found: Factor I (*self-esteem*) is a broad evaluative dimension characterized at the positive pole by success and

satisfaction with life's affairs. It corresponds closely to Osgood's evaluative factors, and appears to be the domain implied by most writers when they refer to the self-concept. Factor II (*anxiety-tension*) is defined by feelings of tiredness, emptiness, fear, and confusion. This dimension, more so than Factor I, may reflect symptoms of distress that are amenable to change during therapy. Factor III (*independence*) appears related to subjectively perceived intelligence and leadership ability. The opposite pole may represent passive acceptance, conformity, and felt inadequacy and inability to lead. Factor IV (*estrangement*) is characterized by feelings of distance and artificiality versus a natural and easy relationship with others. Factor V (*body image*) is similar to Osgood's potency dimension. It is defined by adjectives related to body size and physical strength or potency (e.g., small, light, underweight versus large, strong, rugged, heavy).

Two additional factors were obtained but not interpreted.

The results help explain the findings of investigators who have noted poor correspondence between different tests of the self-concept and low correlation between the self-concept and external criteria of adjustment. It seems likely that these investigators dealt with measures which may have unwittingly confounded several self-concept variables. While it is generally agreed that the extremes of thinking, feeling, and behaving described here are typically observed in psychiatric groups, such states are not *qualitatively* different from those found in normals. Thus, whether "normals" or disturbed Ss are dealt with, tests which treat self-ratings in simple additive fashion may obscure attributes or dimensions of the self-concept that should be treated independently.

REFERENCE

- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

(Received August 20, 1959)

¹ An extended report of this study may be obtained without charge from Philip A. Smith, System Development Corporation, 2500 Colorado Avenue, Santa Monica, California, or for a fee from the American Documentation Institute. Order Documentation No. 6116, remitting \$1.75 for microfilm or \$2.50 for photocopies.

² Formerly at the Ann Arbor VA Hospital. This paper reports one of a series of studies on the behavior of psychiatric patients carried out by D. W. Bostion, R. Ging, G. L. Hover, and J. J. Lasky of the Psychiatry Research Committee at that hospital. Funds for processing data on electronic digital computing facilities at the University of Michigan Statistical Laboratory were provided by the research committee. The consultation of E. L. Kelly is also gratefully acknowledged.

THE EFFECTS OF CHLORPROMAZINE ON PSYCHOLOGICAL TEST SCORES¹

ABE J. JUDSON

Utica College of Syracuse University

AND BARBARA W. MACCASLAND

Marcy State Hospital, Marcy, New York

As a consequence of the widespread use of chlorpromazine with new admissions in psychiatric hospitals, patients seen for diagnostic testing have usually had drug medication for several weeks prior to testing. Because the medication obviously introduces a new factor into the testing situation, it is important to evaluate the effect of the drug on test performance. The aim of this investigation is to determine the effects of chlorpromazine on the following tests: the Wechsler-Bellevue, the Rorschach, the Bender-Gestalt, the Porteus Maze, and the MMPI.

Experimental and control groups were selected from new admissions to a state hospital. All Ss were administered the test battery within 4 to 10 days after admission and before any drug medication had been prescribed. Ss were placed on either chlorpromazine or placebo after completion of the battery and retested 30 days later. Both drug and placebo were discontinued for 48 hours prior to retesting to minimize the influence of the drug's side effects on test performance. There were no significant differences between the two groups in age or education and the modal diagnosis in both groups was schizophrenia. Both male and female patients were included in the sample. There were 18 Ss in the drug group, 21 Ss in the control group. The chlorpromazine dosage received by each experimental S varied with his sensitivity to the drug but for most Ss it was 500 mgs. daily.

The results indicate that under the conditions of the study, chlorpromazine has little or no ef-

fect on psychological test scores. On the Wechsler-Bellevue and Bender-Gestalt, there were no significant differences in gains between the two groups. The latter result differs from Winter and Frederickson's (1956) report that the drug impaired Bender-Gestalt performance. However, they attribute their finding to the side effects of the drug and our procedure tended to reduce these effects. On the Porteus Maze, there was some improvement in performance for the control group and virtually no change for the experimental group but the difference in gains falls short of significance ($t = 1.60$). The difference is in the same direction as that reported by Porteus (1957) who found that chlorpromazine reduced Maze Test scores.

The MMPI and Rorschach results are essentially negative. There were no significant differences in gains on any of the MMPI scales. The Rorschach was scored for $W\%$, $D\%$, $d, dd\%$, $F\%$, $F - \%$, $M\%$, $FM\%$, $FC\%$, $CF, C\%$, $A\%$, $H\%$, $ShR\%$, P , R , and RT . Again, there were no significant differences in gains on any of these variables. Although the results are generally negative, it is worth noting that on the F scale of the MMPI and $F - \%$ score of the Rorschach, both measures of unrealistic thinking, the drug group showed greater improvement ($t = 1.83$ and 1.32 , respectively). These results are not significant but they do point to a possible effect of chlorpromazine on test performance.

REFERENCES

- PORTEUS, S. D. Maze test reactions after chlorpromazine. *J. consult. Psychol.*, 1957, 21, 15-21.
WINTER, W. D., & FREDERICKSON, W. F. The short-term effects of chlorpromazine on psychiatric patients. *J. consult. Psychol.*, 1956, 20, 431-434.

(Received August 26, 1959)

¹ An extended report of this study may be obtained without charge from Abe J. Judson, Utica College of Syracuse University, Utica, N. Y., or for a fee from the American Documentation Institute. Order Document No. 6156, remitting \$1.25 for microfilm or \$1.25 for photocopies.

SIMULTANEOUS RECORDING OF BALES AND CHAPPLE INTERACTION MEASURES DURING INITIAL PSYCHIATRIC INTERVIEWS¹

A. PAUL HARE, NANCY WAXLER,

Harvard University

GEORGE SASLOW, AND JOSEPH D. MATARAZZO

Medical School, University of Oregon

The content of standardized initial psychiatric interviews is analyzed through the use of the Bales (1950) categories and interaction profile and correlated with simultaneously recorded Chapple measures of the same interviews (Matarazzo, Saslow, & Matarazzo, 1956). Subjects are 24 adults referred to the outpatient psychiatry department in a large hospital. Their average age is 36; 17 are female and 7 male.

The characteristics of the partially standardized interview are described in detail in the article by Matarazzo et al. (1956). The interviewer, an experienced psychiatrist, applies a behavior stress in the second and fourth periods of the interview. The second period consists of a silence stress as the interviewer fails to respond by waiting up to 15 seconds after the subject has finished speaking. In the fourth period the interviewer interrupts the subject for a total of 12 times. Time limits are set on the length of each period.

Two scores which are comparable to the Chapple scores of Units and Action were derived from the Bales data by dividing the total number of acts for a patient into two components. The number of times a patient spoke, his Contributions, was computed by counting the number of times the patient began speaking to the psychiatrist. This score is similar to the Chapple score for Patient's Units. The total number of Bales

acts for each patient was divided by the number of Contributions to give a measure of Duration, the average number of acts per contribution. This Duration score based on the Bales data is similar to the Chapple score for Patient Action. No measure similar to the Chapple Silence score appeared to be derivable from the Bales data.

The rank order correlations between the major Chapple variables of Units, Action, and Silence, and three scores derived from the Bales data, Total Acts, Contributions and Duration, are given in Table 1.

The correlation between number of Chapple Patient Units and Bales (Patient) Contributions (+.96) is as high as that obtained for interobserver reliability of Chapple Patient Units for observers using two Interaction Chronographs (Phillips, Matarazzo, Matarazzo, & Saslow, 1957). The correlation between Chapple Patient Action and Bales Patient Duration (+.92) is also high enough to make the two measures interchangeable in the present sample.

REFERENCES

- BALES, R. F. *Interaction process analysis*. Cambridge: Addison-Wesley, 1950.
MATARAZZO, J. D., SASLOW, G., & MATARAZZO, RUTH G. The interaction chronograph as an instrument for objective measurement of interaction during interviews. *J. Psychol.*, 1956, 41, 347-367.
PHILLIPS, JEANNE, MATARAZZO, J. D., MATARAZZO, RUTH G., & SASLOW, G. Observer reliability of interaction patterns during interviews. *J. consult. Psychol.*, 1957, 21, 269-275.

(Received August 29, 1959)

TABLE 1
RANK ORDER CORRELATIONS BETWEEN BALES AND CHAPPLE VARIABLES
(N = 24)

	2	3	4	5	6
1 Units (Chapple)	-.98	-.14	-.42	+.96	-.93
2 Action (Chapple)		-.05	+.50	-.90	+.92
3 Silence Chapple)			-.35	+.01	+.17
3 Total Acts (Bales)				-.47	+.68
5 Contributions (Bales)					-.91
6 Duration (Bales)					

¹ An extended report of this study may be obtained without charge from A. Paul Hare, Dept. Social Relations, Harvard University, Cambridge, Mass., or for a fee from the American Documentation Institute. Order Document No. 6154, remitting \$1.25 for microfilm, or \$1.25 for photocopies.

SOCIAL-PSYCHOLOGICAL FACTORS AND THE PSYCHIATRIC COMPLAINTS OF DISTURBED CHILDREN¹

RAYMOND G. HUNT,

Washington University

JACK L. ROACH,

New York State Department of Social Welfare

AND ORVILLE GURRLIN

Youth Board of Buffalo

This study explored relations between symptomatic complaints made by parents about children in outpatient treatment and social status, sex, religion, age, family size, ordinal position, parental age differences, parent-child age differences and family status (presence or absence of both parents).

Data were collected from records of 265 cases in an urban child guidance clinic. Complaints were ordered to seven commonly used categories of disorder: Habit (e.g., enuresis), Personality (e.g., depression), Neurotic (e.g., tics), Conduct (e.g., stealing), Somatic (e.g., asthma), Educational (e.g., poor grades), and Mental deficiency. Frequencies of complaints as a function of the factors under study were analyzed in contingency tables.

Habit disorders were found to be significantly more frequent in younger children and tended to be so among females and among only and oldest children, but were unrelated to any of the other factors. Personality disorders exhibited no relationships to any of the factors under study nor did somatic complaints. Except for a tendency to be more frequent in "normal" as compared with "broken" homes, the same was true of neurotic disorders.

¹ An extended report of this study may be obtained without charge from R. G. Hunt, Psychology Dept., Washington University, St. Louis 5, Mo. or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 6155, remitting \$1.75 for microfilm or \$2.50 for photocopies.

Conduct disorders were significantly more frequent in children from "broken" homes and tended to be so in older children, only children and those with three or more siblings. They were unrelated to the other factors. Educational disabilities were related only to sex, being significantly more frequent in males.

Complaints of mental deficiency were related to most of the factors studied. They were more frequent among intermediate and youngest children and children from "normal" homes. Prevalence was also greater among children farthest in age from their parents and tended to be more frequent among Catholics.

To shed further light on the age relationship here a chi square analysis revealed complaints of mental deficiency to be more frequent among children whose mothers had been 30 years or older at the child's birth. This, of course, is consistent with the well known tendency for the incidence of mental deficiency to increase with maternal age at conception.

Finally, children from "nonnormal" families were significantly overrepresented in the clinic and were significantly older than "normal" family children.

These modest results suggest several possibilities, one of which is that social factors are less important in shaping disorders of children than of older groups. It is also possible though that intake policies and/or cultural attitudes toward psychiatry constitute "filters" producing spuriously homogeneous patient populations.

(Received September 25, 1959)

MARITAL

SATISFACTION AND PARENT CONCEPTS

RENAUD MANN LOCKE

Research Station, State University of Illinois

Although on a very different rationale, both Freudians and social psychologists generally hold the process of parental identification to be an important factor in adult sexual adjustment. Research into the phenomenon of self-identification with the parent of the same sex is uncovering an increasing number of factors relating to the process (Beier & Ratzenburg, 1953; Ewing, 1954; Ferguson, 1941; Fordyce, 1953; Seward, 1945; Sopchak, 1952; Winch, 1951). Whether and how such identification is related to marital interaction is important to an understanding of the marriage relationship and needs to be explored empirically. A similar factor which may relate to the dynamics of the marital relationship is the degree to which a marriage partner equates his spouse with his parent of the opposite sex. Studies (Ferguson, 1938; Hamilton, 1929; Terman, 1938) previously investigating the association have been inconclusive. Studies in mate selection (Commins, 1932; Mangus, 1936; Strauss, 1946a, 1946b, 1947; Winch, 1958) have generally found that a positive relationship exists between the image one has of his parent of the opposite sex and the spouse he selects.

This investigation has sought to determine whether identification of self with parent of the same sex and the equation of spouse with parent of the opposite sex is associated with marital satisfaction. It has investigated two hypotheses which may be stated in the null form: *There is no difference in population means between two groups defined as satisfactorily and less satisfactorily married with regard to the degree of congruence:*

I. *Between the subject's self-concept and his concept of the parent of the same sex;*

II. *Of the concepts the subject holds of his spouse and of his parent of the opposite sex.*

METHOD

The Criterion and Sample

"Satisfactorily" and "less satisfactorily" married were operationally defined by 22 items found by Locke (1951, 1959) to discriminate between successfully married and divorced subjects. These items were subsequently used by Wallace (1947) and by Swan (1952) and found valid and reliable. The Locke items include inquiry into the permanence of the marriage, adaptability, interests, consensus, affection, sex satisfaction, and general satisfaction. Each item is scored according to weightings determined by Locke's study, and a total score is determined by adding these. As a second, independent measure the Terman (1938) self-rating happiness scale was observed to substantiate the findings. The Terman scale is a horizontal line which runs from "Extremely happy" through seven indicated degrees of marital bliss to "Extremely unhappy." Terman assigned weights to each position: 15, 12, 9, 6, 3, 0, and 0. These weightings were used in this study.

The original population consisted of 594 former students who had been enrolled in Family Life classes, 1948-50, at the University of Minnesota, and who in a 1953 survey indicated that they were married. These 241 men and 353 women were sent 26 classification questions: age, sex, vocation, income, parental background, etc., and the Modified Marital Adjustment Scale as suggested by Locke (1951) plus the Terman self-rating happiness scale. Responses were received from 454 persons (76.5% of the group).

The Locke items were scored and ranked. The distribution ranged from 118 to 52 points: median 102, mean 101.13, and standard deviation of 10.71. Using points Q_3 , score 109, and Q_1 , score 96, as cutting points, those with the highest scores, mean 111.94, and those with the lowest, mean 85.94, were selected to participate in the second step of the study.

Requests were sent to the 116 individuals in the upper group and 108 in the lower group soliciting the participation of their spouses. Code numbers

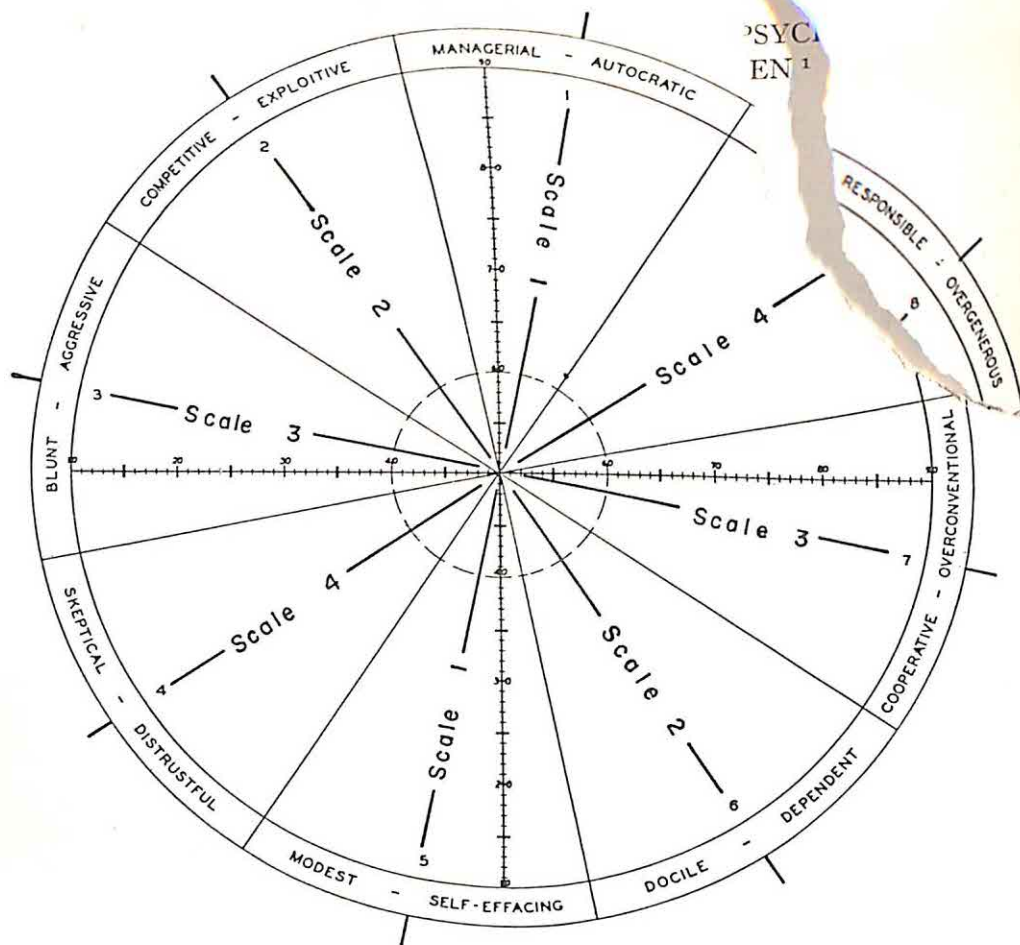


FIG. 1. Interpersonal check list profile illustrating personality variables according to octants and scales.

were used and confidentiality assured. In the upper group 71.5% of the spouses returned Locke questionnaires which they had been requested to mark; 60.5% of the lower group spouses returned them. The fact that fewer in the lower group were willing to participate may be indicative of their lack of marital adjustment (Wallin, 1942). Three couples in the lower group indicated they would not participate because they were either separated or in the process of divorce; one other low scoring couple was divorced and remarried in the nine months' interval between the beginning and end of data collection. No such incidents were discovered in the upper group.

When the Locke scores were computed for all spouses returning questionnaires, and the mean values of the mates of the upper quarter subjects compared with the lower quarter, the difference was significant beyond the .01 level. From these respondents the two final groups, S couples, satisfied, and U couples, less satisfied, were selected after the arbitrary decision to consider as satisfied those couples *both* of whose scores were above the Q_3 point, and as less satisfied

those below the Q_1 point, on the original distribution. Forty couples were thus designated for the U group of couples and 41 for the S.

The happiness scale was observed to determine the degree of differentiation between the spouses of the S and U couples; the means of the two groups were tested for differences by the t test and found significant beyond the .01 level. It can be concluded with confidence that the two samples were significantly different with regard to degree of marital satisfaction as tested by the instruments of this study.

Measurement of Self and Other Concepts

The 41 spouses of the original members of the study scoring above Q_3 and the 40 of those scoring below Q_1 were mailed copies of the Interpersonal Check List (LaForge, 1955; Leary, 1956) and each was directed to mark it without the knowledge or cooperation of the spouse. When this check list was returned, one was mailed to the other spouse with identical directions and request. In both samples there was a 100% response and replies were returned from both individuals.

The two groups were then examined for homogeneity by means of information gained through the 26 items of descriptive personal data. Out of 29 tests of significance by means of chi square, one difference between the groups was found. This falls well within the limits of chance and the populations may be considered as drawn from the same population. The factors examined were: age, age at marriage, number of years' education, number of years married, number of children, children's ages, vocational status, income, number of working wives and wives' income, religious affiliation, present residence, birthplace, community of childhood, vocation of parents, marital status and education of parents. The one difference found was that the fathers of the husbands of the S' Couples had a greater number of years' education.

The Interpersonal Check List, was chosen as the self-concept instrument. It is a list of 128 descriptive self-referent items selected to represent interpersonal variables, each made up of 8 items. These variables are arranged in a circular profile and combined into descriptive octants in such a way that the opposite octant represents an opposing variable; i.e., each intersecting line on the graph may be thought of as a continuum with the subject's score being located according to the number of items checked within that specific category. For the purpose of this study the two octants making up the one continuum have been designated as a scale, of which there are four for each concept measured. Scale 1 is made up of variables termed Managerial-Autocratic, Octant AP, at one end of the continuum, and Modest-Self effacing, Octant HI, at the other. Scale 2, Competitive-Exploitive, BC, vs. Docile-Dependent, JK; Scale 3, Blunt-Aggressive, DE, vs. Cooperative-Over conventional, LM; Scale 4, Skeptical-Distrustful, FG, vs. Responsible-Over generous, NO. See Figure 1.

Some of the items from the check list are: "well thought of," "apologetic," "somewhat snobbish," "clinging vine," "outspoken," "will confide in anyone," "often gloomy," and "encourages others."

The check list is scored by octants, the criterion being the number of items checked. When an octant is considered as one part of a continuum with an opposing octant as the other part, a single score on the continuum may be obtained by subtracting the upper score from the lower, thus locating a point on the continuum. Four such points may be determined for each individual concept marked on the check list.

Instruction to subjects marking the ICL solicit direct, conscious-level, verbal responses which tap the subject's perception and his willingness to express this perception. Such an overt, conscious self-report is appropriate to the phenomenological assumption that the phenomenal field coincides with subjective awareness. Operationally, the self and other concepts may be defined for the purpose of this study in terms of all the statements an individual makes about himself or his world as they are represented by these check list items. Each subject marked four concepts: self, spouse, mother, and father. On each

of these four concepts the four scale scores were determined. It is these four scale scores which are used as a basis for comparing the degree of congruity of one personality with another: self with parent of the same sex; spouse with parent of the opposite sex. The term "identification" is thus defined operationally for this study as the measured degree of agreement between the self-concept and the concept of an "other." The process of consciously ascribing similarities or differences in terms of the ICL to various family members is termed "equating." If the subject describes his wife in the same way that he describes his mother, he would be said to equate them.

A discrepancy score between each personality on each scale was determined. The couples groups were divided on the basis of sex so that sex differences might be observed. Distributions of the discrepancy scores were made. A transformation of the data was then made in order to make the distributions more nearly normal and to equalize the variance. In this way the assumptions of the *t* test of significance were met. The variable of discrepancy determined from scores on the scales may be designated X. The new variable of discrepancy was taken and called Y and the transformation was effected: $Y = \log (X + 1)$.

Such an analysis provides a way of discovering whether pairs of concepts are more congruent in the satisfied groups than in the less satisfied, but it does not provide any way to discover the direction or trend of these differences in terms of descriptive personality traits. The two previous studies (Hamilton, 1929; Terman, 1938) have been concerned only with whether there were or were not similarities in concepts between spouse and parent of the opposite sex and what the relation was to marital satisfaction. Terman asked if spouse and parent of the opposite sex were "alike in general physical type." Hamilton's subjects were asked if their spouses and parents of the opposite sex were "similar physically" and "similar mentally." Because the ICL is conveniently organized in terms of eight related personality variables, it is possible to define discrepancies not only by their magnitude but in terms of whether one concept has more or less of the qualities described by Scales 1, 2, 3, and 4.

In order to do this, the mean raw score of each ICL octant was determined for each sex, both S and U groups, on each concept: self, spouse, parent of the same sex, and parent of the opposite sex. Scale scores were then obtained thus:

$$X \text{ Scale 1} = \bar{X}_{AP} - \bar{X}_{HI}$$

$$X \text{ Scale 2} = \bar{X}_{BC} - \bar{X}_{JK}$$

$$X \text{ Scale 3} = \bar{X}_{DE} - \bar{X}_{LM}$$

$$X \text{ Scale 4} = \bar{X}_{FG} - \bar{X}_{NO}$$

These scale scores were then plotted on a continuum representing the two octants composing each scale. Comparisons of the concepts of self and parent of the same sex, and spouse and parent of the opposite sex, could thus be observed. The distance between

TABLE 1

SIGNIFICANT DIFFERENCES BETWEEN LESS SATISFIED AND SATISFIED GROUPS WHEN DISCREPANCY BETWEEN SELF AND PARENT OF THE SAME SEX IS TESTED

Discrepancy Between Wife's Self-Concept and Her Concept of Her Mother					Discrepancy Between Husband's Self-Concept and His Concept of His Father				
ICL Scale	\bar{U}_x	\bar{S}_x	t	Level of Significance	ICL Scale	\bar{U}_x	\bar{S}_x	t	Level of Significance
1	.735	.585	2.24	.05	1	.731	.550	2.55	.01
2	.539	.484	.83	accept	2	.558	.424	2.20	.05
3	.545	.536	.15	accept	3	.609	.550	.82	accept
4	.601	.599	.03	accept	4	.670	.468	2.89	.01

concept scores represents the magnitude of the discrepancy. The raw score values themselves provide the comparisons indicating the *direction of differences* between the concepts in terms of the ICL personality variables on the four scales (see Figures 2 and 3).

RESULTS

When Hypothesis I which deals with the phenomenon of identification of self and parent of the same sex was examined, it was obvious that there was a sex difference. Women in the satisfactorily married group were *not* found to identify with their mothers to a greater degree than those in less satisfactory marriages. The null hypothesis was accepted on three of the four scales. On every scale the mean discrepancy scores were found to be higher in the U group, but were not statistically greater except on Scale 1. Men in the S group, however, were found to have lower mean discrepancy scores on every scale and significantly lower ones on Scales 1, 2, and 4. One may conclude that men who are satisfied with their mar-

riages as indicated by this study perceive themselves and their fathers as being similar to a greater degree than do men who are less satisfied.

The second null hypothesis, which held that there is no difference in population means between two groups defined as satisfactorily and less satisfactorily married with regard to the degree of congruence of the concepts the subject holds of his spouse and of his parent of the opposite sex, was rejected on six of the eight scales (see Table 2). Again a sex difference was found. Wives in the S couples were found significantly different on every scale from wives in the U couples. It can be concluded that women who rated high in marital satisfaction perceived their husbands and their fathers as similar to a greater extent than those who scored low on the marital satisfaction scale.

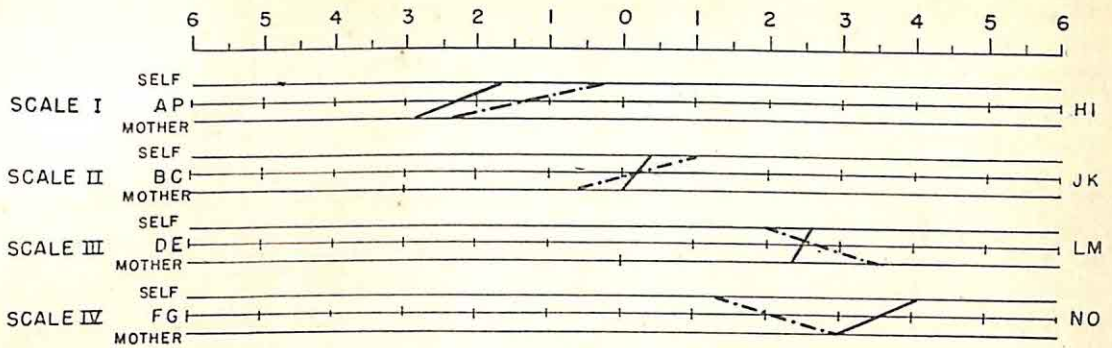
When comparisons were made between the husbands of the S and U Couples, the findings were not so conclusive. Husbands who were more satisfied were found to have a sig-

TABLE 2

SIGNIFICANT DIFFERENCES BETWEEN LESS SATISFIED AND SATISFIED GROUPS WHEN DISCREPANCY BETWEEN SPOUSE AND PARENT OF THE OPPOSITE SEX IS TESTED

Discrepancy Between Wife's Concept of Her Husband and Her Concept of Her Father					Discrepancy Between Husband's Concept of His Wife and His Concept of His Mother				
ICL Scale	\bar{U}_x	\bar{S}_x	t	Level of Significance	ICL Scale	\bar{U}_x	\bar{S}_x	t	Level of Significance
1	.644	.515	1.77	.05	1	.636	.521	1.46	accept
2	.599	.488	1.79	.05	2	.608	.545	.89	accept
3	.809	.577	3.41	.01	3	.665	.536	1.90	.05
4	.709	.497	2.86	.01	4	.703	.534	2.56	.01

WIFE: SELF AND MOTHER



HUSBAND: SELF AND FATHER

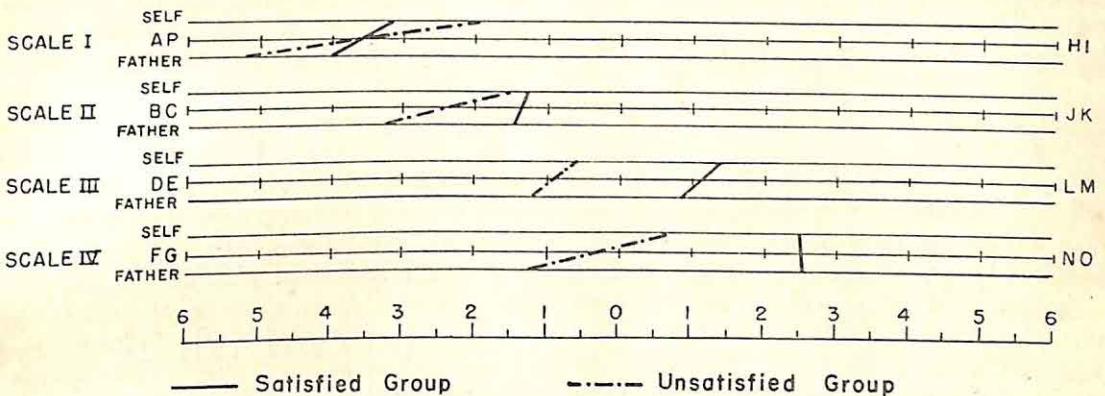


FIG. 2. Directional trends of scores of self and parent of the same sex when compared on four interpersonal check list scales.

nificantly greater congruence of concepts of wife and mother on Scales 3 and 4. Although the mean discrepancy scores on Scales 1 and 2 were lower in the S group, the difference was not significant. The question remains whether the husband's perception of wife and mother is associated with satisfaction in marriage.

Directional Differences

When directional differences were observed as plotted on the continua of the four scales, the following conclusions could be drawn with regard to the comparisons made between the S and U groups when self and parent of the same sex was observed (see Figure 2).

The directional differences which exist between women and men when they compared themselves with their parent of the same sex were most obvious on Scales 3 and 4 on which U women saw themselves as being less co-

operative and responsible than their mothers, but U men saw themselves as having more of these qualities than do their fathers. All S subjects, men and women, viewed themselves in relation to the parent of the same sex in the same directions.

When wives compared themselves with their mothers on Scale 1, both S and U wives thought themselves less managerial and autocratic than they thought their mothers; however, U women thought themselves *considerably* less so.

Scale 2 differences were not significant between S and U women and the directional trend for both groups is the same with the subject seeing self as more docile and dependent than she sees her mother.

Scale 3 differences which were not great enough to be statistically significant are, however, in different directions. S subjects perceive themselves and mothers as very similar,

WIFE: HUSBAND AND FATHER

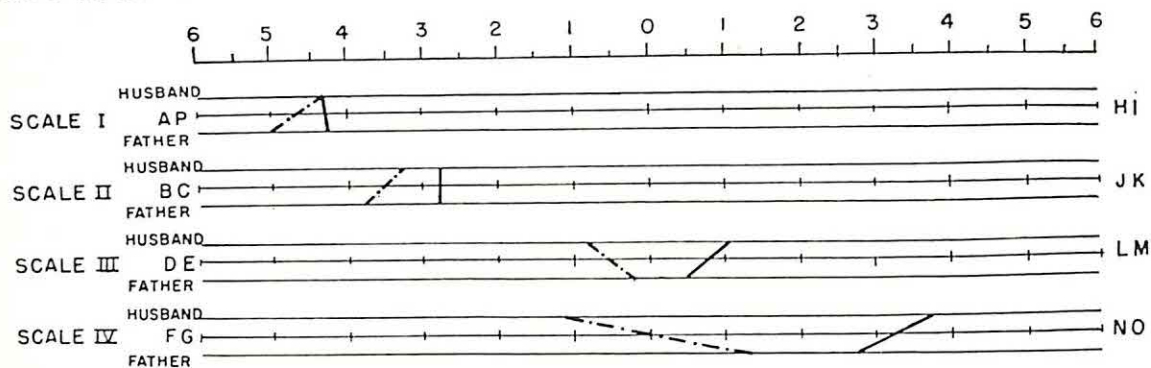


FIG. 3. Directional trends of scores of spouse and parent of the opposite sex when compared on four interpersonal check list scales.

with *selves* being a bit more cooperative and conventional. U women see *mothers* as being much more cooperative than they see themselves.

On Scale 4 the difference in discrepancy between the concepts for the two groups of women was not significant. Each group perceived the mother similarly, but self-perceptions were very different. U women thought themselves much less responsible and generous than S women thought themselves or than U women thought their mothers.

S and U women agree on all four scales more often on their perception of their mothers than they do on self-perceptions. This was not true for men subjects. When men compared selves and father, S and U men tended to see themselves as more like each other than they saw their respective fathers as alike.

On Scales 1 and 2 the direction is the same for both S and U groups of men, but the difference between self and father is significantly greater in the U group. U men saw their fathers as much more managerial and competitive than S men saw theirs.

Scale 3 discrepancies between self and father were not significant; U men saw both selves and fathers as more blunt and aggressive than S men saw themselves and fathers.

The discrepancy between self and father for the two groups of men was significant on Scale 4. S husbands saw themselves and fathers similarly and considered themselves as cooperative and generous. U men, on the other hand, saw themselves as being low in this quality and saw their fathers as being even much lower; so that the fathers' score is a positive score in the FG octant, indicating skepticism and distrust.

Graphs indicating the directional trends in which S and U subjects compared their spouses with their parents of the opposite sex indicate that on every scale S and U women saw spouses in relation to their fathers in the opposite direction (see Figure 3).

Both S and U women saw husbands as moderately managerial and autocratic, but U wives saw their fathers as being more so.

On Scale 2, U women perceived their fathers in comparison with husbands as being slightly more competitive, S women saw them as being slightly less.

When the wives' perceptions of S and U husbands were compared on Scale 3, S husbands were seen as more cooperative than U husbands, and S fathers were seen *less* cooperative than S husbands. Not so with U fathers who were seen as *more* cooperative than the husbands of U wives.

The directional trends on Scale 4 are like those on Scale 3 only more so. S and U husbands are seen much differently by their wives, with S wives seeing husbands as very responsible and generous and U wives seeing their husbands in the opposite direction, as skeptical and distrusting. Fathers of both groups of women were perceived as being responsible and generous, but S wives thought fathers had less of this quality than their husbands had, and U wives thought their fathers had more of it than husbands.

When men compared wives and mothers, discrepancies between these were less than the discrepancies found among women subjects when they compared husbands and fathers; the directional trends were also not so great nor in so clear a pattern.

S husbands saw wives as a little less managerial and autocratic than their mothers. U husbands perceived the two in the opposite way; wives were more "bossy."

Very little difference exists in the two groups on Scale 2.

Both S and U husbands saw wives as more cooperative than mothers; S mothers were seen as more cooperative than U mothers were.

On Scale 4 the differences go in opposite directions for the S and U groups. Men who were in satisfactory marriages saw both their wives and mothers as being more responsible

and generous than U men saw their women-folk. S men thought their wives and mothers were quite similar in this respect but mothers were somewhat more generous; U men judged their wives to be the more generous.

DISCUSSION

The answer to the question of why men who identify with fathers were found more often in satisfactory marriages than in unsatisfactory ones, but women who identify with mothers were not, may lie within the nature of modern American culture and the confused role of women. A modern wife's conception of herself as similar to her mother as measured by this test is unrelated to her satisfaction in marriage. Doing things the way mother did, seeing self as a female of mother's cut, apparently is not a factor closely associated with the satisfaction a modern wife finds in marriage. Elements other than her allegiance to or expectation of the role of wife and mother as she saw it portrayed in the previous generation are evidently of greater import in her marriage.

The same conclusions cannot be drawn from the data with regard to men. The findings may reflect a greater stability of the husband-father role in our changing culture and suggest that those men who see themselves fulfilling the husband role as they saw their fathers fulfilling it are more satisfied in their own marriages. Satisfaction in marriage for the male may be associated with his own feelings of masculinity which have been learned from the father; or may characterize men who, in terms of Freudian theory, have identified with the male or father-image.

All subjects thought selves less managerial, autocratic, competitive, and exploitive than their parent of the same sex, but all S subjects thought this difference not so great. That parents are seen as being high in these dominant qualities may reflect an attitude that children, even adult children, hold in regard to parents—that they are pretty dictatorial people.

If the mother serves as model of self for the woman, one would conclude that the U wife might feel a sense of disappointment in herself because she saw herself as being less cooperative, generous, and responsible than

her mother. The S wife who perceived more of these qualities in herself than in her mother may see herself as an improvement over the mother.

If lack of modesty, self-effacement, docility, and dependence is considered as a lack of masculine strength, men in the U group, when comparing selves with fathers, might also consider themselves as lesser men than their fathers whom they saw as having fewer of these "weak" characteristics. The greatest difference between the S and U men, which is on Scale 4, indicates that U men saw themselves and their fathers as considerably more distrustful and skeptical than S men saw themselves and fathers. The fathers of U husbands were perceived as being even *more* skeptical than were the men themselves.

The sex differences found with regard to equating spouse and parent in the second hypothesis is of even greater interest. Women in satisfactory marriages saw their husbands and fathers as significantly more alike on all four scales than women in less satisfactory marriages. The findings were not so definite in regard to the association of satisfaction and the congruence of the husband's images of wife and mother; two scales were significantly and positively related, two were not.

These findings are not in agreement with those of the Hamilton study (1929) and Ferguson's (1938) interpretation, but Terman's study (1938) which found happy wives preferred husbands and fathers "not of the opposite type" would seem compatible with the results of this study. This kind of sex difference lends support to the assumption that in satisfying marriages there is a relationship between wife and husband that reflects the relationship between parent and child in that the wife sees her husband as "father" and he, too, sees himself as "father." Wives can continue comfortably in a dependent role relating to the husband much as they did to father; not so, the husband who would like to continue in a filial relationship with his wife. This is a cultural expectation that may account for the lack of concisive evidence that marital satisfaction is associated with the husband's equating of wife and mother.

Observation of the directional trends when

wives compared husbands and fathers would indicate that U wives thought their husbands more modest, self-effacing, docile, and dependent than their fathers (Scales 1 and 2). Husbands seen in this way might be considered not so masculine as fathers. In keeping with this generally less favorable picture of their spouses, U wives thought husbands more blunt and distrustful than fathers. S wives equated husbands and fathers rather closely on Scales 1, 2, and 3, but thought husbands more responsible and generous than fathers on Scale 4. If women look to their fathers as models for husbands, the U wife might well feel disillusioned in her husband, while the S wife, who saw the two similarly except for increased generosity in her husband, might be especially well pleased with the man she had married.

The men saw their wives and mothers statistically more discrepant only on Scales 3 and 4. On Scale 3 the direction in which the S and U husbands viewed their wives and mothers was similar; both thought wives a bit more cooperative and conventional. The fact that U men saw both wives and mothers as less responsible and generous than S men saw theirs may indicate that a generally less favorable impression of the female is held by men in less satisfactory marriages; mothers of men in the U group were seen as even less generous than their wives. It may also indicate these U men actually have their lives peopled with less "helpful, big-hearted, kind, and reassuring" women.

Although the premises of this study are based in perceptual differences and the ICL responses cannot claim to represent objective reality, one cannot escape the conjecture that perhaps not only the processes of identification and equation are associated with marital satisfaction, but that, indeed, the S and U subjects have real personality differences. In observing Figures 2 and 3 one readily notes that the differences between all S and U subjects are consistently greatest on Scale 4 in every comparison. The U subjects, male and female, place themselves, spouses, and parents farther into the "Skeptical-Distrustful" octant than do the S subjects. The perceptions of selves and others held by U subjects indicate more of the quality measured by

FG items such as: "often gloomy, bitter, resentful, complaining, jealous, stubborn." One may speculate that persons dissatisfied in marriage may actually be more skeptical and distrustful, may have married skeptical and distrustful spouses, and have been reared by skeptical and distrustful parents.

It also may be conjectured that the U women tended to see children generally as less socially desirable persons than parents. This conclusion is drawn from the similarity with which U women saw themselves in relation to their mothers and husbands in relation to their fathers. In each case sons and daughters were seen as generally less "forceful, dominating, confident, independent," and less "cooperative, friendly, warm, tender, considerate," than parents.

SUMMARY

Two hypotheses dealing with the relationship of parent image to marital satisfaction were investigated. Two groups of married couples, $N = 41$ and 40 , defined as "satisfactorily" and "less satisfactorily" married, indicated their perception of self, spouse, parent of the same sex, and parent of the opposite sex on the Interpersonal Check List. Discrepancy scores were determined between the concepts (a) of self and parent of the same sex and (b) of spouse and parent of the opposite sex and were tested for significance of differences by means of the t test.

Directional trends in the differences between concepts were graphed on continua representing each of the four ICL scales and were observed in order to discover in what ways the concepts held by the S and U groups differed in terms of the eight personality variables represented by the ICL.

Men in satisfactory marriages were found to identify with their fathers to a significantly greater degree on three out of four of the scales. On each of the four scales unsatisfactorily married men saw their fathers as being more dominant and less loving than themselves. Such differences were not found among the two women's groups; on three of the scales the discrepancies between self and mother were not found statistically different for S and U women. When directional trends were observed, U women were found to see

themselves as less cooperative and responsible than their mothers while S women thought themselves to have slightly more of these qualities than their mothers.

When the S and U groups were compared regarding the degree of congruence the subject held of his spouse and of his parent of the opposite sex, a significant relationship was found on all four scales of the women's scores, but was not clear with regard to men. Women in satisfactory marriages perceived their husbands as similar to their fathers to a greater degree than did less satisfied wives.

When the nature of these perceptions was observed by individual ICL scales, dissatisfied women were found to perceive their husbands as less dominant and competitive and more blunt, aggressive, skeptical, and distrustful than they saw their fathers. The satisfied wives who found the opposite on each of the four scales might well perceive their husbands as being more desirable men than their fathers.

Men in satisfactory marriages as contrasted with those in less satisfactory marriages perceived their wives as similar to mothers to a significantly greater degree on Scales 3 and 4; on the other two scales there appeared to be no relationship. For this reason, the findings are considered inconclusive. An analysis of the nature of these differences indicated no difference in direction on Scale 3; however, on Scale 4, U men saw spouses as less generous than mothers, but S men saw them as slightly more. U husbands saw both their mothers and their wives as being reliably more skeptical and distrustful than did husbands in the satisfied group.

Congruence in perception of self and parent of the same sex was found associated with marital satisfaction for men. Equation of spouse and parent of the opposite sex was significantly related to marital satisfaction for women; the findings for men were not conclusive.

REFERENCES

- BEIER, E. G., & RATZENBURG, F. The parental identification of male and female college students. *J. abnorm. soc. Psychol.*, 1953, 48, 569-572.
- COMMINS, W. D. The marriage age of oldest sons. *J. soc. Psychol.*, 1932, 3, 487-490.

- EWING, T. M. Changes in attitude during counseling. *J. counsel. Psychol.*, 1954, 1, 232-239.
- FERGUSON, L. W. Correlates of marital happiness. *J. Psychol.*, 1938, 6, 284-294.
- FERGUSON, L. W. The cultural genesis of masculinity-femininity. *Psychol. Bull.*, 1941, 38, 584-585 (Abstract).
- FORDYCE, W. E. Applications of a scale of dependency to concepts of self, ideal self, mother, and father. Unpublished doctoral dissertation, Univer. of Washington, 1953, *Dissertation Abstr.* 1953, 13, 584-585.
- HAMILTON, G. V. *A research in marriage*. New York: Boni, 1929.
- LA FORGE, R., & SUCZEK, R. An interpersonal check list. *J. Pers.*, Sept. 1955, 24, No. 1, 94-112.
- LEARY, T. Kaiser Foundation Psychology Research. *Multilevel measurement of interpersonal behavior*. (Manual). Berkeley, Cal.: Psychological Consultation Service, 1956.
- LOCKE, H. J. *Predicting adjustment in marriage*. New York: Holt, 1951.
- LOCKE, H. J., & WALLACE, K. M. Short marital-adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living*, August 1959, 21, 251-256.
- MANGUS, A. R. Relationships between the young woman's conception of her intimate male associates and of ideal husband. *J. soc. Psychol.*, 1936, 7, 403-420.
- SEWARD, GEORGENE. Cultural conflict and the feminine role: An experimental study. *J. soc. Psychol.*, 1945, 22, 177-194.
- SOPCHAK, A. L. Parental identification and 'tendency toward disorders' as measured by the MMPI. *J. abnorm. soc. Psychol.*, 1952, 47, 159-165.
- STRAUSS, A. The ideal and the chosen mate. *Amer. J. Sociol.*, 1946, 52, 204-208. (a)
- STRAUSS, A. The influence of parent-image upon marital choice. *Amer. Sociol. Rev.*, 1946, 11, 554-559. (b)
- STRAUSS, A. Personality needs and marital choice. *Soc. Forces*, 1947, 25, 332-335.
- SWAN, R. The application of a couple analysis to the MMPI in marriage counseling. Unpublished doctoral dissertation, Univer. of Minnesota, 1953.
- TERMAN, L. M. *Psychological factors in marital happiness*. New York: McGraw-Hill, 1938.
- WALLACE, K. M. Construction and validation of marital adjustment and prediction scales. Unpublished doctoral dissertation, Univer. of Southern California, 1947.
- WALLIN, P. The characteristics of participants in a social psychological study. Unpublished doctoral dissertation, Univer. of Chicago, 1942.
- WINCH, R. F. Further data and observations on the oedipus hypothesis: The consequence of an inadequate hypothesis. *Amer. sociol. Rev.*, 1951, 16, 784-795.
- WINCH, R. F. *Mate selection*. New York: Harper, 1958.

(Received May 15, 1959)

NEED SATISFACTION AND EDWARDS PPS SCORES IN MARRIED COUPLES¹

IRWIN KATZ, SAM GLUCKSBERG, AND ROBERT KRAUSS²

New York University

What relationships exist between the personalities of marriage partners? Winch (1958) found in a study of 25 married couples that the needs of spouses tended to be dissimilar, which he took as evidence that mates select one another on the basis of complementarity. His results are in conflict with the data of several earlier studies, in which low positive correlations between spouses' traits were sometimes obtained, but never negative correlations. For example, Burgess and Wallin (1953) gave 31 items of Thurstone's Neurotic Inventory to 1,000 engaged couples and found on 14 items significantly greater similarity than resulted from random matchings, while no items yielded reliably negative correlations.

Winch (1958) attributed the disagreement between his own data and the earlier results to the relative superficiality of the self-inventories that were used by previous investigators, as compared with his own reliance on depth interviews and the TAT. However, Winch's scoring method contained a contaminating feature that casts doubt on the validity of his findings. Each judge rated all of the characteristics of a given subject, so that scores on different traits cannot be regarded as independent of one another. Of 71 significantly negative correlations between like traits and significantly positive correlations

between unlike traits (388 correlations were run in all) an indeterminate number could actually have been spurious reflections of the raters' implicit theories of trait organization.

To get new evidence on husband-wife personality relationships we employed a modified form of the Edwards Personal Preference Schedule (EPPS). The Edwards test is not highly susceptible to response faking of the social desirability type (Edwards, 1957) and provides reasonably independent scores on different motivational variables, as indicated by generally low intercorrelations (Edwards, 1953). In addition, self-ratings and predicted self-ratings of need satisfaction in the marriage were obtained, so that degree of satisfaction could be compared with the similarity-dissimilarity of mates' scores on the Edwards variables.

METHOD

Subjects

Fifty-six volunteer couples were obtained through the cooperation of psychology graduate students at New York University. The mean length of time married was 5 years, with a range of 6 months to 22 years. The mean number of children was .7, range 0 to 3. For husbands the mean age was 29.6 years, range 23 to 52; and for wives the mean age was 28 years, range 20 to 46. Every person was white, with a background of at least 12 years of schooling. All couples were middle class, as determined by the husband's occupation or educational status.

Testing Procedures

All Ss filled out a shortened version of the EPPS in which the number of personality variables had been reduced from 15 to 11 for reasons of economy. By eliminating Order, Intracception, Change, and Heterosexuality, we reduced the total number of items from 225 to 110. Ss were given a second questionnaire which required ratings of the extent to which the spouse tended to satisfy or thwart each of the 11 needs that were measured by the EPPS. Needs

¹ This investigation was supported in part by a research grant (M-2421) from the National Institute of Mental Health, Public Health Service.

The present study makes use of a shortened version of the EPPS. Since there is evidence to indicate that item responses obtained to selected items isolated from the context of the EPPS are not comparable to those obtained within the context, the results of this research cannot be considered applicable to the standardized complete form of the EPPS.

² Now at Bell Laboratories.

were defined by the statements associated with them in the EPPS. Under each definition was a seven-point graphic rating scale on which Ss were to respond "independently of the importance of the need in you; to the extent that you have this need at all, does your spouse's behavior tend to satisfy it or thwart it or not affect it one way or the other." Ss were given another copy of the rating sheet with instructions to predict the spouse's responses.

RESULTS AND DISCUSSION

Comparison of Mates' EPPS Scores

Winch found that husband-wife similarity on need variables was significantly lower than that of randomly matched men and women, as measured by *Q* type correlations. We computed rho correlations between EPPS scores on 11 variables for all couples. The mean correlation was .24, with a *SD* of .37. For 25 randomly matched men and women the mean value of rho was .26, and the *SD* was .25. Thus, similarity between mates' need profiles was close to random expectation.

Husbands and wives were also compared on scores on particular needs. Winch had obtained five times as many complementary as noncomplementary relationships. We selected 21 combinations of EPPS variables which seemed relevant to the hypothesis of complementarity, as follows:

Husband	Wife	
Abasement	Abasement	—
Abasement	Dominance	+
Abasement	Aggression	+
Achievement	Achievement	—
Affiliation	Affiliation	—
Aggression	Aggression	—
Aggression	Abasement	+
Autonomy	Autonomy	—
Autonomy	Nurturance	—
Deference	Dominance	+
Deference	Deference	—
Dominance	Deference	+
Dominance	Abasement	+
Dominance	Dominance	—
Endurance	Endurance	—
Exhibition	Exhibition	—
Nurturance	Autonomy	—
Nurturance	Nurturance	—
Nurturance	Succorance	+
Succorance	Succorance	—
Surrrorance	Nurturance	+

Pluses and minuses show the type of relationship indicated by the complementarity hypothesis. Many of these pairings had been

tested by Winch. Since we were interested only in ascertaining whether reliably positive or negative relationships existed, a chi square test of association between high-low EPPS scores for husbands and high-low EPPS scores for wives was performed for each need pair. The dichotomies were based upon median splits of the distributions of need scores in each sex. Four pairings yielded *p* values < .05 and one pairing yielded a *p* < .10. Four pairs of like needs were positively related:

Abasement-Abasement ($\chi^2 = 5.42, p < .02$)
 Affiliation-Affiliation ($\chi^2 = 2.81, p < .10$)
 Autonomy-Autonomy ($\chi^2 = 9.07, p < .01$)
 Nurturance-Nurturance ($\chi^2 = 4.54, p < .05$)

One pair of unlike needs were negatively related:

Husb. Succorance-Wife Nurturance ($\chi^2 = 3.87, p < .05$)

Thus, in every instance the direction of association was *opposite* to that predicted by Winch's complementarity hypothesis.

Total Need Satisfaction and Relationships between Spouses' EPPS Scores

For each S, self-ratings of satisfaction on all 11 needs were added together to yield a total satisfaction score, after which separate distributions of scores for men and for women were set up. Then High Satisfaction and Low Satisfaction groups were established by selecting from each distribution the 20 highest and the 20 lowest individuals. Association between High-Low Satisfaction and similarity-dissimilarity of spouses' EPPS scores was tested via chi square for men and women separately.

In women three notable relationships were found between total satisfaction and degree of resemblance of own and spouse's EPPS scores. High Satisfaction wives, as compared with Low Satisfaction wives, showed a trend toward less similarity to their husbands on Aggression ($\chi^2 = 2.84, p < .10$). On Nurturance and Succorance respectively, High Satisfaction wives were significantly more similar to their husbands than were Low Satisfaction wives (in both cases $\chi^2 = 4.86, p < .05$).

Thus, for wives complementarity was not generally related directly to total satisfaction.

Among husbands, three significant findings and one trend revealed in every case greater complementarity in the High Satisfaction group than in the Low Satisfaction group. In the following, interspousal similarity was less for High Satisfaction men in the first three cases, and greater in the last one.

Achievement-Achievement ($\chi^2 = 5.65, p < .02$)
 Succorance-Succorance ($\chi^2 = 4.26, p < .05$)
 Dominance-Dominance ($\chi^2 = 3.28, p < .10$)
 Husb. Succorance-Wife Nurturance ($\chi^2 = 4.86, p < .05$)

Summarizing, it would appear that the nature of husband-wife need relationships, as measured by the EPPS, is different for various need pairings, for different degrees of total need satisfaction, and for the sexes.

Total Need Satisfaction and Individual EPPS Scores

Chi square tests were used to ascertain whether there was any relation between total need satisfaction (20 highs vs. 20 lows) and a person's EPPS scores (high vs. low by median split). Among wives two significant findings indicated positive relationships between total satisfaction and two EPPS variables:

Nurturance ($\chi^2 = 3.94, p < .05$)
 Succorance ($\chi^2 = 6.46, p < .02$)

Total satisfaction of wives was positively related to two EPPS scores of their husbands (one trend and one significant finding):

Nurturance ($\chi^2 = 2.80, p < .10$)
 Achievement ($\chi^2 = 6.46, p < .02$)

Also, there were trends toward negative association between wives' total satisfaction and two EPPS scores of their husbands:

Abasement ($\chi^2 = 2.86, p < .10$)
 Autonomy ($\chi^2 = 3.50, p < .10$)

Comparing husbands' total satisfaction with their own EPPS scores revealed two negative relationships, one a trend and the other significant:

Affiliation ($\chi^2 = 8.12, p < .01$)
 Dominance ($\chi^2 = 3.64, p < .10$)

Total satisfaction of husbands showed four significant relationships with EPPS scores of their wives. Two were positive:

Succorance ($\chi^2 = 4.91, p < .05$)
 Nurturance ($\chi^2 = 6.46, p < .02$)

Two were negative, as was a trend:

Autonomy ($\chi^2 = 4.91, p < .05$)
 Dominance ($\chi^2 = 4.91, p < .05$)
 Aggression ($\chi^2 = 3.28, p < .10$)

It can be seen that for both men and women, total need satisfaction in marriage was related to some of one's own and one's spouse's EPPS scores. Interestingly, satisfaction of both wives and husbands was positively related to wives' scores on Nurturance and Succorance.

Additional Comparisons

The product-moment correlation between husbands' and wives' total satisfaction scores was .55, indicating a fair degree of agreement. Men's predictions of their spouses' satisfaction ratings were significantly lower than their own satisfaction ratings (mean difference = 2.97, $CR = 3.09, p < .01$), and significantly lower than their spouses' predictions for them (mean difference = 2.82, $CR = 2.24, p < .02$). The mean self-ratings on satisfaction of husbands and of wives were virtually the same, as were women's predictions of their husbands' satisfaction ratings and their own ratings.

Chi square tests of association (high-low vs. high-low) yielded no p values below .20 when length of marriage was compared with each of the following: interspousal need similarity, interspousal similarity of total satisfaction scores, total satisfaction scores of either men or women, accuracy of predicted satisfaction ratings by either men or women. Similarly, there was no tendency toward a relationship between total satisfaction of either sex and interspousal need similarity.

SUMMARY

Fifty-six married couples were administered a modified version of the EPPS and a questionnaire for measuring degree of need gratification afforded by the spouse. One aim of the study was to examine the hypothesis of inter-

spousal complementarity of needs by comparing the EPPS scores of mates. Also, relations between total satisfaction and EPPS scores were explored. The findings were:

Interspousal correlations on 11 EPPS variables were close to the random level.

Twenty-one husband-wife pairings of like and unlike EPPS variables produced five relationships contradictory to the complementarity hypothesis, and none supportive of it.

Degree of total satisfaction of wives was not consistently related to interspousal need complementarity. But total satisfaction of husbands was positively associated with interspousal complementarity in four need pairings.

Thus, the nature of husband-wife need relationships, as measured by the EPPS, was different for various need pairings, for different degrees of total satisfaction, and for the sexes.

Among wives, total satisfaction was positively related to own EPPS scores on Nurture and Succorance, and was related to certain EPPS scores of their spouses. Hus-

bands' total satisfaction tended to be related to some of their own EPPS scores, and to some of their spouses' scores—including positive relationships with spouses' Succorance and Nurture scores.

Husbands' predictions of their spouses' total satisfaction scores were lower than their own ratings and lower than their spouses' predictions for them. The mean self-ratings of husbands and wives were virtually the same.

There were no relationships between length of marriage and other variables or between total satisfaction and interspousal need similarity.

REFERENCES

- BURGESS, E. W., & WALLIN, P. *Engagement and marriage*. Philadelphia: Lippincott, 1953.
EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corp., 1953.
EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
WINCH, R. F. *Mate-selection*. New York: Harper, 1958.

(Received May 22, 1959)

OBJECTIVE MEASURES OF PERCEPTION IN SCHIZOPHRENICS AND NORMALS¹

RUTH COOPER

Teachers College, Columbia University

Many views have been proposed concerning the nature of impairment in schizophrenia. Some theorists postulate disturbances in cognitive functioning as one of the primary factors in the schizophrenic disorder (Angyal, 1946; Bleuler, 1950; Goldstein, 1943). Others suggest that schizophrenia is primarily an affective disorder (Freud, 1947; Sullivan, 1953). Despite differences in orientation and emphasis, most theories agree that impaired reality testing is a central phenomenon in schizophrenic illness.

While it may be true that schizophrenic impairment involves cognitive difficulties, or emotional conflicts, or both, another possible hypothesis is that schizophrenia involves, more simply and directly, impaired perceptual functioning.

Klein (1951) averred that perception was perhaps the most important area of investigation of individual organization because it was the individual's point of contact with reality. Since poor reality testing is considered a primary symptom of schizophrenia, and perception is intimately related to how an individual deals with his environment, one would expect less accurate perception by people whose reality testing mechanisms are inadequate.

Recent research indicates a growing interest in the use of psychophysical measures as means of differentiating between clinical

groups. Studies on autokinetic effect (Voth, 1947), size constancy (Lovinger, 1956; Rausch, 1952-53), and judgments of weights (Salzinger, 1957) support the notion that objective, psychophysical measures may well provide useful diagnostic tools (Stevens, 1951). Their particular attractiveness lies in the premise that effects of more complex cognitive and emotional factors are minimized when simple, objective techniques are used.

The present study was designed to investigate differences of response between schizophrenics and normals to three visual stimuli like those frequently used in psychophysical experiments. The general hypotheses were that schizophrenics are less accurate and more variable than normals in judging objective visual stimuli. Moreover, since there are varying degrees of schizophrenic disturbance, from mild to severe, it was further hypothesized that degree of perceptual accuracy and variability are related to degree of disturbance.

METHOD

Subjects

Schizophrenic group. Sixty-six (66) male veterans were selected from psychiatric wards of Manhattan and Northport Veterans Administration Hospital. They had been diagnosed schizophrenic, were without organic brain pathology, and had not received shock therapy within a month of testing. Of this group, 46 Ss had been hospitalized for less, and 20 for more than 24 months. Included were paranoid, undifferentiated, hebephrenic, catatonic, and simple schizophrenics. Only 3 had been diagnosed as acutely schizophrenic. The group was representative of present day psychiatric veteran patients who are, for the most part, chronic, process-type schizophrenics.

Normal group. Thirty-six (36) male veterans were selected from general medical and surgical wards of Manhattan Veterans Hospital. Patients with past psychiatric referral or neurological involvement were excluded. The sample included 26 ambulatory, minor

¹This paper is adapted from a dissertation submitted to Teachers College, Columbia University in partial fulfillment of the requirements for the PhD degree. The author wishes to express sincerest appreciation to her sponsor, Joel R. Davitz, whose advice and encouragement were of inestimable value in the execution of this investigation. She wishes also to thank Laurance F. Shaffer, Edward J. Shoben, Jr., and Gertrude P. Driscoll for their constructive and helpful criticism.

postsurgery patients, and 10 long-term tuberculosis patients, some confined to bed.

The groups were comparable with respect to age and educational level. Visual acuity was within normal limits; those *Ss* who ordinarily wore glasses used them during the testing situation.

Instruments and Administration

S was informed that a research project was being conducted at the hospital to determine some of the factors that influence perception. He was told that his active cooperation would be appreciated, despite the fact that the findings might not be of immediate practical benefit to him. The purpose of such instructions was to minimize any anxiety *S* might have felt if he thought test results might be used "against" him in any way. This approach seemed effective in virtually every case. *Ss* seemed motivated and interested throughout the testing situation.

Prior to actual testing, *S* was required to reproduce an extremely simple block design, the model for which *E* constructed. The purpose was to determine whether *S* could follow instructions. Only one *S* had to be excluded for inability to follow instructions.

All three tasks were administered during one session, with Galton Bar judgments the first task for all *Ss*. To control for possible effects of fatigue or waning interest, order on the succeeding tasks was varied; half the *Ss* performed Sander Parallelogram judgments next; the other half, the Kunnapas Squares judgments.

Galton Bar. An adaptation of the Galton Bar was constructed. It consisted of a wooden board (20" wide \times 18" tall) onto which were mounted two wooden slide rules from which all markings had been removed on the side facing the *Ss*. The entire instrument, except for the areas behind the moveable part of the slide rule, was painted black and kept scrupulously free of any markings which might provide clues or anchoring points for the *Ss*. The back of the slide rule (the moveable part) was marked off on a millimeter scale, and ranged from 0 to 200 mm. Either side (right or left) could and did serve as the constant or the variable. The length of the constant was kept at 150 mm.

E instructed *S* to indicate by saying "Stop" when the variable (which *E* manipulated at all times) appeared equal to the constant. A total of 16 trials was given: 4 with the constant on the right; 4 with constant on the left; 4 more with constant on the right, and, finally, 4 with constant on the left. Initial variable settings were either much longer or much shorter than the constant. Order for the size settings was longer-shorter-shorter-longer, shorter-longer-longer-shorter.

Sander Parallelogram. A model of the Sander Parallelogram was constructed of stiff gray cardboard, and was mounted on a wooden frame. Except for the left diagonal, which had to be moveable, the parallelogram was drawn on the cardboard in black India ink. The width of all lines, including the thin strip of metal which served as the moveable diagonal,

was 4 mm. The length of the fixed diagonal was 150 mm. The variable covered a range from 71 mm. to 156 mm. The back of the parallelogram was marked off by means of Polar Grid paper, from which readings were easily converted into millimeters.

Administration was the same as for the Galton Bar, except that there were 8 trials instead of 16, and constant was always on the same side because of the nature of the parallelogram. As with the Galton Bar, *S* was instructed to indicate by saying "Stop" when the variable appeared equal in length to the constant.

Kunnapas Squares. Two sets of squares, adapted from a figure-ground instrument devised by Kunnapas (1955), were cut from white oaktag. Each set contained 11 cards; the first comprised squares whose side (like that of the constant) measured 9 cm.; the side of the second set measured 21 cm. A black line was inscribed in the center of each square. The range of line lengths was 11 mm., i.e., from 45 mm. to 55 mm. All lines were 4 mm. wide, and the standard was 50 mm. long.

The squares were presented on a large, light blue background and were always 13 cm. apart, lying flat on the blue background. The constant was kept on the right side for one ascending (45 mm. to 55 mm.) and one descending series of presentations, and on the left for the second ascending and descending series. This procedure was followed for both the 9 and 21 cm. sets, providing a total of 44 judgments for each set.

S was instructed to judge whether the variable line was "longer" or "shorter" than the constant. Since each set contained one square whose inscribed line was exactly equal to the constant, *E* instructed *S* to give his best guess whenever he experienced real difficulty in making a particular judgment. *S* was not informed that any of the lines were, in fact, equal to the constant.

Montrose Rating Scale. Each psychiatric patient was rated for severity of illness by the ward psychiatrist. If *S* was not well known to the psychiatrist, the ward psychologist or nurse was asked to do the rating. The Montrose Rating Scale consists of seven five-point rating scales intended to measure *S's* functioning in seven areas—reality testing, emotionality, communication, human relationships, aspirations, manifest overt behavior, and intellectual functioning. A split-half reliability check computed in the present study yielded an *r* of .87.

Treatment of Data

Judgments on each instrument were recorded without regard for sign, in terms of distance from objective equality. Thus, each judgment was treated as a difference score (i.e., an error), and average error and variance were computed for each *S* and for each group.

On the Kunnapas Squares task, errors were weighted according to distance from equality. Thus, judging a 45 mm. line as "longer" than the 50 mm. standard was assigned a weight of 5, whereas judging a 49 mm. line as "longer" carried a weight of 1. The measure

of variability was the number of reversals from "shorter" to "longer" to "shorter."

Additional analysis. Because of the apparent similarity of the perceptual tests, the data were analyzed to discover whether any relationship existed between tests with respect to accuracy and to variability of performance.

1. The *t* test was used to test for significance of differences of scores between Normal and Schizophrenic groups.

2. The *F* test was computed for significance of differences in variability between Normal and Schizophrenic groups.

3. A product moment correlation was computed between scores on each of the three tests for the Normal group.

4. *Phi* coefficients were computed between:

- severity of illness and scores on each of the tests;
- severity of illness and variability on each of the tests;
- variability on each test for each group of Ss;
- scores on each test for the Schizophrenic group.

5. Split-half reliabilities were computed for each instrument and the Spearman-Brown correction was applied.

RESULTS AND CONCLUSIONS

Results are presented for each hypothesis after a brief statement of the hypotheses.

Accuracy of Judgments

The first hypothesis stated that normals are more accurate than schizophrenics in judging objective visual stimuli. Differences between

TABLE 1

MEAN ERROR SCORES OF THE SCHIZOPHRENIC GROUP AND THE NORMAL GROUP IN JUDGMENTS ON THE GALTON BAR, THE SANDER PARALLELOGRAM AND THE 9 AND 21 CM. KUNNAPAS SQUARES

Instrument	Group		<i>t</i>	<i>p</i>
	Schizophrenic	Normal		
Galton Bar	8.78	4.43	3.50	.005
Sander Parallelogram	24.07	18.63	5.33	.0005
Kunnapas Square, 9 cm.	14.26	4.11	17.8	.0005
Kunnapas Square, 21 cm.	15.33	4.14	6.99	.0005

TABLE 2

MEAN VARIABILITY SCORES OF THE SCHIZOPHRENIC AND THE NORMAL GROUP IN JUDGMENTS ON THE GALTON BAR, THE SANDER PARALLELOGRAM, AND THE 9 AND 21 CM. KUNNAPAS SQUARES

Group	Galton Bar	Sander Parallelogram	Kunnapas Squares	
			9 cm.	21 cm.
Schizophrenic	48.50	30.89	10.17	8.95
Normal	12.44	12.62	2.45	.31

Note.—All differences significant at or beyond .01 level.

group means were tested with the *t* test and found significant at or beyond the .005 level on each instrument. The first hypothesis, therefore, was supported. Table 1 presents the results.

Variability of Judgments

The second hypothesis, that normals are less variable than schizophrenics when judging objective visual stimuli, was also supported. The *F* test yielded results significant at or beyond the .01 level. Results are summarized in Table 2.

Relationship of Severity of Illness to Accuracy and Variability of Judgments

The data did not support the hypotheses of relationship between severity of illness and accuracy and variability of judgment. Correlations with psychiatrists' ratings of severity of illness were essentially zero.

TABLE 3

INTERCORRELATIONS OF THE SCHIZOPHRENIC Ss' SCORES ON THE GALTON BAR, THE SANDER PARALLELOGRAM, AND THE 9 AND 21 CM. KUNNAPAS SQUARES

	Sander Parallelogram	Kunnapas Squares	
		9 cm.	21 cm.
Galton Bar	.36*	.15	.21
Sander Parallelogram		.21	.36*
Kunnapas Squares, 9 cm.			.52*

* Significant at or beyond .01 level.

TABLE 4

INTERCORRELATIONS OF THE NORMAL Ss' SCORES ON THE GALTON BAR, THE SANDER PARALLELOGRAM, AND THE 9 AND 21 CM. KUNNAPAS SQUARES

	Sander Parallelogram	Kunnapas Squares	
		9 cm.	21 cm.
Galton Bar	.34*	-.22	-.04
Sander Parallelogram		.06	-.28
Kunnapas Squares, 9 cm.			-.02

* Significant at .05 level.

TABLE 5

INTERCORRELATIONS OF THE SCHIZOPHRENICS Ss' VARIABILITY ON THE GALTON BAR, THE SANDER PARALLELOGRAM, AND THE 9 AND 21 CM. KUNNAPAS SQUARES

	Sander Parallelogram	Kunnapas Squares	
		9 cm.	21 cm.
Galton Bar	.21	.23	.00
Sander Parallelogram		.00	.03
Kunnapas Squares, 9 cm.			.44*

* Significant at .001 level.

Relationship Between Performance on the Tests

To test for relationship between performance with respect to accuracy, Pearson's r was computed for normals, and phi for schizophrenics.

Table 3 demonstrates that correlations, significant at or beyond the .01 level, were obtained between scores on the Galton Bar and Sander Parallelogram, Sander Parallelogram and 21 cm. Kunnapas Squares, and 9 and 21 cm. Kunnapas Squares by the schizophrenics. No other correlations achieved significance. Thus, the Sander Parallelogram accounts for two correlations in accuracy which, though low, are significant.

For the normals, as Table 4 shows, a significant correlation was found in accuracy of responses between the Galton Bar and Sander

TABLE 6

INTERCORRELATIONS OF THE NORMAL Ss' VARIABILITY ON THE GALTON BAR, THE SANDER PARALLELOGRAM, AND THE 9 AND 21 CM. KUNNAPAS SQUARES

	Sander Parallelogram	Kunnapas Squares	
		9 cm.	21 cm.
Galton Bar	.91	.65	.89
Sander Parallelogram		.80	.89
Kunnapas Squares, 9 cm.			.95

Note.—All correlations significant at .0001 level.

Parallelogram. No other correlations achieved significance.

Tables 5 and 6 indicate that with respect to variability no significant correlations were obtained for the schizophrenics except between the two subtests of the Kunnapas Squares. The normals, on the other hand, achieved correlations significant at the .0001 level between all of the tests.

DISCUSSION

Differences between Groups

The hypotheses that schizophrenics are less accurate and more variable than normals in judging objective visual stimuli were supported. Since both cognitive complexity and emotional factors were kept to a minimum, one might speculate that schizophrenia is neither merely a cognitive nor an emotional disturbance, but that it also involves perceptual impairment which is not readily accounted for by either cognitive or emotional theories of schizophrenia. It is particularly significant that such impairment can be demonstrated in simple, objective, readily quantifiable tasks.

Relationship of Severity of Illness to Accuracy and Variability of Judgment

The hypotheses that there are relationships between severity of illness and accuracy and variability of judgment were not supported. Several possibilities present themselves to account for the results. One is that many raters were involved in the study, and it was not

TABLE 7
RELIABILITY OF MEASURES FOR THE SCHIZOPHRENIC AND NORMAL GROUPS

Test	<i>r</i>	Schizophrenic		Normal	
		Spearman Brown	<i>r</i>	Spearman Brown	correction
Montrose Rating Scale	.77	.87*			
Galton Bar	.48	.65*	.30	.46*	
Sander Parallelogram	.94	.97*	.96	.98*	
Kunnapas Squares, 9 cm.	.60	.75*	.05	.10	
Kunnapas Squares, 21 cm.	.61	.76*	.71	.83*	

* Significant at or beyond .001 level.

possible to check interjudge reliability. An alternative explanation is the possibility that in schizophrenia perceptual impairment has an all-or-none quality. One further possibility is the fact that since really regressed, back ward schizophrenics were excluded from the study, those Ss who were included did, in fact, vary very little in degree of schizophrenic illness.

Intercorrelation among Measures

Small but significant correlations in accuracy scores were obtained for schizophrenics between the Galton Bar and Sander Parallelogram, the Sander Parallelogram and 21 cm. Kunnapas Squares, and the two subtests of the Kunnapas Squares. With respect to variability, not only are schizophrenics more variable than normals, but they also vary more in their variability.

For the normals, the only significant correlation in accuracy scores was between the Galton Bar and Sander Parallelogram. Probably the low reliability of the 9 cm. and the limited range of scores of the 21 cm. Kunnapas Squares account for this finding.

Implications for Future Research

The instruments used in the present research have demonstrated a potential for differentiating between normals and schizophrenics. The relatively small overlap between the two groups' distribution of scores supports the view that the Galton Bar and Sander Parallelogram, in particular, may prove to be worthwhile additions to test batteries now used to differentiate between normals and schizophrenics.

The utility of simple, objective, psychophysical perceptual tasks in differentiating

between clinical groups has been demonstrated. It is not unreasonable to expect that such tests may also be of considerable value in differentiating among personality types. They may, as a result, contribute some fruitful hypotheses for investigation to personality and perception theorists.

SUMMARY

The present study investigated differences in accuracy and variability of perceptual judgments between schizophrenics and normals, using simple, objective, visual stimuli like those used in psychophysical experiments. Significant differences both in accuracy and in variability were obtained between groups.

No relationship was demonstrated between degree of schizophrenic illness and performance on tests.

Correlations in accuracy of performance were demonstrated between some of the measures. With respect to variability, no correlations were demonstrated for the schizophrenics, whereas normals showed highly significant correlations between all tests.

Implications of these findings, both for theory and for future research, were discussed.

REFERENCES

- ANGYAL, A. Disturbances of thinking in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Los Angeles: Univer. California Press, 1946. Pp. 115-123.
- BLEULER, E. *Dementia praecox*. New York: International Universities Press, 1950.
- FREUD, S. *The ego and the id*. London: Hogarth Press, 1947.
- GOLDSTEIN, K. The significance of psychological research in schizophrenia. *J. nerv. ment. Dis.*, 1943, 97, 261-279.

- KLEIN, G. S. The personal world through perception. In R. R. Blake & C. V. Ramsey (Eds.), *Perception*. New York: Ronald, 1951. Pp. 328-355.
- KUNNAPAS, T. M. Influence of frame length on apparent length of line. *J. exp. Psychol.*, 1955, 50, 168-170.
- LOVINGER, E. Perceptual contact with reality in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, 52, 87-91.
- RAUSCH, H. L. Perceptual constancy in schizophrenia: I. Size constancy. *J. Pers.*, 1952-53, 21, 176-187.
- SALZINGER, K. Shift in judgment of weights as a function of anchoring stimuli and instructions in early schizophrenics and normals. *J. abnorm. soc. Psychol.*, 1957, 55, 43-49.
- STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 28-33.
- SULLIVAN, H. S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.
- VOTH, A. C. An experimental study of mental patients through the autokinetic effect. *Amer. J. Psychiat.*, 1947, 103, 793-805.

(Received June 1, 1959)

EGO STRENGTH AND TYPE OF DEFENSIVE BEHAVIOR¹

GERALD F. KING AND MARVIN SCHILLER

Michigan State University

In a study of the construct validity of the Barron (1953) Ego Strength (*Es*) Scale, Schiller (1958) failed to find a significant relationship between this scale and the amount of defensiveness displayed by individuals in a stressful, real life situation. In a postmortem consideration of the negative results, the concept of defensiveness was given some scrutiny. As employed in this and other studies, defensiveness is a quantitative extension of the psychoanalytic typology of defenses. It is a composite or global concept. This paper presents a reanalysis of Schiller's data by exploring the feasibility of using specific types of defenses as a frame of reference instead of the general concept of defensiveness.

In the reanalysis, the initial task was seen as a conceptual one, that of deriving the expected relationship between ego strength and different defensive behavior. While such a relationship is not explicitly outlined in psychoanalytic theory, the parallel between ego development and the developmental sequence of the defense mechanisms seems to offer a basis for making predictions. Blum (1953) and others have indicated that there are a number of ambiguous aspects in the theoretical treatment of defense mechanisms,² but most of these elements can be bypassed in the present considerations.

As an exploratory study, the concern was with the more gross features of psychoanalysis. For example, it is sufficient to utilize the

differentiation that denial and projection are more primitive defense mechanisms than rationalization, more primitive in the sense that they develop and operate earlier in the individual's life. The sequence of defenses that are employed by the child parallels the maturation of the ego, especially in its increasing awareness and adaptation to reality. Both denial and projection are incompatible with reality. As ego functions become more extended, denial and projection become less appropriate mechanisms. Bellak (1958) states,

By and large the earlier it emerges, the more primitive a defense, and the more pathological in adult life. Extensive denial and very marked projection probably vie for first place with regard to seriousness of pathology, by virtue of their effect on the individual's adaptation to reality (p. 20).

Reality is given definite recognition in the defense mechanism of rationalization. According to Fenichel (1954), this mechanism serves as a means of justifying certain behavior by making it reasonable in the light of reality. Thus, rationalization emerges at a higher level of ego development or "ego strength."

To complete the conceptual picture, it seemed reasonable to assume that the relationship between ego maturation and the development of various defenses in childhood has relevance for adult behavior, i.e., ego strength is related to the developmental level of the defense mechanism that is employed. The specific hypothesis was that, in a situation conducive to the elicitation of defensive behavior, level of ego strength is positively related to the relatively greater use of rationalization than either denial or projection.

METHOD

Subjects and Setting

The Ss were 60 male drivers who, due to an excessive number of traffic violations and/or accidents, had been summoned to the Driver and Vehicle Serv-

¹ Based on a master's thesis submitted by the junior author and supervised by the senior author. Grateful acknowledgment is due Terrence M. Allen and Frank Restle for their interest and suggestions.

² Some of the theoretical ambiguities of psychoanalysis in regard to defense mechanisms are as follows: (a) vagueness in the criteria as to what is classifiable as a defense mechanism, (b) lack of specificity in the delineation of their chronological development, and (c) disagreement on whether some of the mechanisms of childhood also occur in normal adults.

ices, Office of the Secretary of State (Michigan), for a re-examination interview to determine their future driving privileges. The ages of the Ss ranged from 17 to 62 years ($M = 25.67$, $SD = 9.04$), with level of education extending from 8 to 16 years ($M = 11.82$, $SD = 1.99$). The Ss were aware that the re-examination interview could result in the loss of their driving permits. Such decisions were made by the interviewers in more than 50% of the cases. When the importance of a driver's license to the average individual is considered, it can be assumed that there were strong elements of anxiety in this real life situation, one that could be expected to elicit defensive behavior.

Procedure

Prior to the re-examination interview, the Ss were individually administered a battery of tests that included the *Es* scale, the WAIS Vocabulary Subtest, and the DDB Inventory (King, 1957).

Measures of Defensive Behavior

The DDB Inventory was designed to measure the orientations of problem drivers toward their past driving records (violations and accidents). In content, it consists of 40 defensive and nondefensive statements that had been culled from the interview remarks of several hundred problem drivers. The 28 defensive statements were selected by judges to reflect "defensiveness against accepting personal responsibility for one's traffic record." The Ss indicated their degree of agreement with the inventory items on a three-point scale. The sum of the weighted responses to the defensive items gives a general defensive score. Computed for several samples, the split-half reliability coefficients for the general defensive score were of the order of .80. In studies of concurrent validity, correlations of the order of .60 were obtained between the DDB measure of defensiveness and interview ratings of defensiveness.

To meet the purposes of the present analysis, the 28 defensive statements were submitted to 10 judges (advanced graduate students), who were asked to use the following categories in classifying the statements according to the type of defense mechanism that was represented: denial, projection, rationalization, and indeterminate. From the items on which there was agreement by at least 8 out of the 10 judges, 4 were selected for each of the three defenses. (More than four items were available for two of the defenses.)

Examples of statements illustrating the different types of defenses are as follows: My record may look bad, but I really don't drive that way (denial); I think that the police are too strict in enforcing the traffic laws (projection); I believe that since a good driver knows how to handle himself, it doesn't hurt to go over the speed limit once in a while (rationalization). Making delineations among the defensive statements in this manner yielded separate "quantitative-typology" scores for denial (*D*), projection (*P*), and rationalization (*R*).

RESULTS AND DISCUSSION

As a preliminary step, the defensive scores (*D*, *P*, and *R*), along with six levels of ego strength from the *Es* scale, were treated with analysis of variance. Table 1 shows that, in addition to the significant difference for between defenses ($F = 32.81$, $p < .01$), the interaction between the *Es* levels and the different defenses was of marginal significance ($F = 1.94$, $p < .05$). The latter finding is compatible with the hypothesis. Using corrections for main effects, i.e., differences between defenses, the scores were plotted, with *Es* being the ordinate and the defensive scores falling on the abscissa. In line with the hypothesis, the expectation was that *R* would be represented by an ascending diagonal, which would be crossed by the descending diagonals of *D* and *P*. The results were moderately good for *R* and *P* but not for *D*.

A correlational analysis provided the strongest support for the hypothesis. First, it was found that *Es* was not significantly correlated with the individual defensive scores: *Es* vs. *D*, $-.10$; *Es* vs. *P*, $-.19$; *Es* vs. *R*, $-.08$. What might be called a measure of the "relative use of rationalization (*RUR*)" was then derived from the following simple formula: $RUR = R - (D + P) + 10$, the constant being added to eliminate negative values. Support was given to the hypothesis by the resulting significant correlation of .27 between *Es* and *RUR* ($p < .05$). No noticeable departure from rectilinearity was suggested by

TABLE 1
ANALYSIS OF VARIANCE OF THE DIFFERENT DEFENSIVE
SCORES WITH SIX LEVELS OF EGO STRENGTH (*Es*)

Source of Variation	df	Sum of Squares	F	p
Between <i>Es</i> Levels	5	9.37	0.89	
Between Ss	54	113.71		
Total Between Ss	59	123.08		
Between Defenses	2	49.24	32.81	<.01
<i>Es</i> × <i>D</i>	10	14.55		
Pooled Ss × <i>D</i>	108	80.77	1.94	<.05
Total Within Ss	120	144.56		
Total	179	267.64		

an inspection of the scattergram. Further analysis revealed that the correlation between *Es* and *RUR* was independent of intellectual level, as measured by the WAIS vocabulary.

In deriving specific defensive scores from the DDB Inventory, only 12 of the 28 defensive items were used. Consideration was given to the possibility that the relationship between *Es* and *RUR* was a function of the 12 items that were selected, that comparable results could be obtained with these items using the general concept of defensiveness. This possibility was not supported when *Es* was correlated with a general defensive score based on the 12 items, i.e., the sum of *D*, *P*, and *R*. The *r* of $-.15$ was in the direction predicted by Schiller (1958) but not significant.

A second sample of *Ss* ($N = 50$), tested in the same setting, was available for a cross-validation of the obtained correlation between *Es* and *RUR*. While the *Ss* in this sample underwent a shorter test battery, the principal difference was that the DDB Inventory was slightly modified. The inventory items were arranged differently, and some were changed in wording. As in the first sample, *Es* was not significantly correlated with the individual defensive scores or the revised general defensive score: *Es* vs. *D*, $-.14$; *Es* vs. *P*, $-.17$; *Es* vs. *R*, $.04$; *Es* vs. *D + P + R*, $-.12$. The correlation between *Es* and *RUR* was again significant ($r = .31$, $p < .05$).³

For the purpose of comparison, the analysis of the same data in accordance with two different orientations can be outlined as follows. In the original treatment of the data (Schiller, 1958), the correlations between level of ego strength and general defensiveness in two independent samples were $-.18$ and $-.20$, both in the predicted direction but not significant. Guided more by the specifics of psychoanalytic theory, the present study yielded small

but significant positive correlations (*rs* of $.27$ and $.31$) between level of ego strength and the relative use of rationalization in the two samples. While the differences produced by the two analyses are by no means striking, the results do suggest that, in the present context, the psychoanalytic typology of defenses provides a more fruitful frame of reference than the general concept of defensiveness.

SUMMARY

Preliminary research employing the general concept of defensiveness failed to find a relationship between level of ego strength, as measured by the Barron (1953) Ego Strength Scale, and the amount of defensiveness shown by individuals in a stressful real life situation. In a reanalysis of the data, the present study explored the possibilities of a different orientation, that offered by the psychoanalytic typology of defense mechanisms. The specific defenses considered were denial, projection, and rationalization. The following hypothesis was suggested by psychoanalytic theory: In a situation conducive to the elicitation of defensive behavior, level of ego strength is positively related to the relatively greater use of rationalization than either denial or projection. The results obtained from two samples supported the hypothesis.

REFERENCES

- BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327-333.
- BELLAK, L. The schizophrenic syndrome. In L. Bellak (Ed.), *Schizophrenia: A review of the syndrome*. New York: Logos Press, 1958. Pp. 1-63.
- BLUM, G. S. *Psychoanalytic theories of personality*. New York: McGraw-Hill, 1953.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- KING, G. F. A scale for measuring "defensiveness" in problem drivers. East Lansing: Highway Traffic Safety Center, Michigan State Univ., 1957.
- KING, G. F., & SCHILLER, M. Note on ego-strength, defensiveness, and acquiescence. *Psychol. Rep.*, 1958, 4, 434.
- SCHILLER, M. Ego-strength, manifest anxiety, and defensiveness. Unpublished master's thesis, Michigan State Univ., 1958.

(Received June 5, 1959)

³ Previous research (King & Schiller, 1958) revealed that the response set "acquiescence" operated in both *Es* and the general defensive score of the DDB Inventory, which suggests that caution should be exercised in the interpretation of correlations obtained with these scales. However, no evidence was found for the presence of acquiescence in *RUR*. The *r* between *Es* and *RUR* was not appreciably affected by attempts to partial out acquiescence.

PERSONALITY CORRELATES OF PEPTIC ULCER PATIENTS¹

SIMONE MARSHALL²

Teachers College, Columbia University

Alexander's theory points out two major aspects in the psychodynamics of peptic ulcer patients: (a) a strong drive for achievement, independence, and self-sufficiency conflicting with (b) persistent infantile wishes to be dependent and protected. The patient remains unaware of his dependency needs and typically describes himself as follows: "I am efficient, active, productive; I give to everybody, help people, assume responsibilities, like to have people depend on me, like to be the effective leader and the self-sufficient, active, or even aggressive personality" (Alexander & French, 1948, p. 110). The Achievement drive and the related personality variables, Efficiency, Responsibility, Dominance, Self-sufficiency and Aggression, are conceived as defensive reactions against unacceptable or "ego-alien" dependency needs.

Other aspects of the defensive pattern are also considered by Alexander and his colleagues. Therese Benedek (1946) and Catherine Bacon (1948) have noted the "severe self-criticism" of the ulcer patient and his need "to lead a very ethical life" (Bacon, 1948, p. 135) as a reaction against his receptive trend. Such an emphasis on self-criticism and *strict moralism* with the consequent *lack of self-acceptance* are likely to be reflected in a large discrepancy between self-concept and ideal self. The guilt and shame experienced in relation to dependent longings are generalized and emotional responses become restricted. This "absence of give and take of emotional responses" as well as the "inability to show and even feel emotion" (Alexander & French, 1948, p. 189) may be termed *emotional inhibition* and is associated with im-

poverished interpersonal relationships. Such a limited ability to interact with others, coupled with the need to maintain a rigorous moralistic code, is not likely to allow much inner life and introspection so that another aspect of the defensive pattern characterizing ulcer patients may be the inability to question one's motives and others' intentions; that is, there is a *lack of insight*. In considering the relationships among variables such as Strict Moralism, Lack of Insight, and Emotional Inhibition in the context of a need for Achievement, it appears that the ulcer patient may attempt to attain his goals by accepting only the more obvious social guidelines and by exercising restraint. In a sense, the emphasis may be on adjustment to explicit cultural sanctions, or in a word—*conformity*. This great reliance on conventional values does not encourage innovations but rather tends to maintain the status quo, since most changes are experienced as threatening to a precarious emotional balance. Thus, *resistance to change* may also be included in the defensive behavior pattern of peptic ulcer cases.

In sum, the personality variables of particular significance in the defensive behavior pattern of peptic ulcer patients include: Efficiency, Responsibility, Dominance, Self-sufficiency, Achievement and Aggression, as specifically emphasized in Alexander's theory. The variables of Strict Moralism, Lack of Self-acceptance, Emotional Inhibition, Lack of Insight, Conformity and Dislike for Change were derived from a review of the cases reported in the literature by Alexander and his followers.

Studies which corroborate Alexander's theoretical position consist mostly of clinical observations made in the course of psychotherapeutic interviews (Van der Heide, 1948; Garma, 1950) and investigations using projective techniques (Modell & Potter, 1949; Brown, Bresnahan, Chalke, Peters, Poser, & Tougas, 1950). While most researchers agree

¹ This paper is based on a dissertation submitted in candidacy for the degree of Doctor of Philosophy at Teachers College, Columbia University. The writer wishes to acknowledge the invaluable assistance of the sponsoring committee, Joel R. Davitz, Gertrude P. Driscoll, and Rosalea A. Schonbar.

² Now with the Ossining Board of Education, Ossining, New York.

with Alexander's assumption concerning the existence of strong dependency needs in peptic ulcer cases, the particular defense mechanisms used to prevent realization of the dependent desires are frequently questioned (Kapp, Rosenbaum, & Romano, 1947; Prince, 1950; Weinberg, 1951). The present study was designed to investigate the pattern of defensive behavior proposed by Alexander as particularly characteristic of the peptic ulcer patient.

HYPOTHESES

General Hypothesis

The general hypothesis tested was that peptic ulcer patients, in comparison to a non-ulcer psychosomatic group and a nonpsychosomatic group, endorse and value achievement, responsibility, and related personality variables.

Experimental Hypotheses

The following research hypotheses were formulated on the basis of the personality variables comprising the defensive pattern proposed by Alexander:

1. Peptic ulcer patients, in comparison to nonulcer psychosomatic and nonpsychosomatic patients, describe themselves as (a) more dominant, (b) more aggressive, (c) more efficient, (d) more responsible, (e) more self-sufficient, (f) more strictly moralistic, (g) more emotionally inhibited, (h) more conforming, (i) more achievement-oriented, (j) less interested in change, (k) less intraceptive.

2. Peptic ulcer patients, in comparison to nonulcer psychosomatic and nonpsychosomatic patients, show greater self-ideal discrepancy.

METHOD

Subjects

The experimental group, composed of 40 patients with X-ray evidence of ulcer, was compared to two control groups. One control group, C₁, included 20 psychosomatic cases other than gastrointestinal disturbances such as asthma and neurodermatitis. The other control group, C₂, was comprised of 40 patients without psychosomatic or psychiatric disorders, i.e., upper respiratory infections and minor surgical recoveries.

All patients were on active duty in the Armed Forces and were hospitalized at U. S. Army Hospital, Fort Dix, New Jersey. Experimental and control groups were equated for age, education, intelligence

TABLE 1
ANALYSES OF VARIANCE APPLIED TO DEMOGRAPHICAL CHARACTERISTICS OF EXPERIMENTAL AND CONTROL GROUPS

Sources of Variation	df	Mean Square	F	p
Age				
Between	2	6.60	.13	ns
Within	97	52.20		
Total	99			
Education				
Between	2	3.70	.94	ns
Within	97	3.92		
Total	99			
Socioeconomic Level				
Between	2	3.85	1.89	ns
Within	97	2.05		
Total	99			
Intelligence ^a				
Between	2	22.36	.09	ns
Within	78	256.79		
Total	80			

^a Only 80% of the GT scores were available. The distribution of nonavailable GT scores was nine in the experimental group, three in the control group C₁, and eight in the control group C₂.

gence (GT score of Army Classification Battery, 1948) and socioeconomic level (Hamburger Classification, 1957) as shown in Table 1.

Measuring Instruments

Two instruments were used to test the hypotheses of the present study: the Edwards Personal Preference Schedule (EPPS), and a specially designed rating scale, the Peptic Ulcer Index, which provides for self and ideal self measures of eight of the variables considered in the defensive pattern against the awareness of dependency. Both instruments were presented to Ss as questionnaires included in a survey of personal attitudes of hospitalized versus nonhospitalized soldiers.

Construction of the Peptic Ulcer Index

A total of 160 statements, 20 for each of the following variables, were compiled. The variables were defined in terms of descriptions given in cases reported by Alexander and his followers:

Dominance: characterized by a person who likes to have people depend on him, likes to be the effective leader, manifests highly competitive tendencies.

Aggression: characterized by occasional outbursts of anger, but more often by a subdued hostility.

Efficiency: characterized by a person who is active, hard-working, capable, and wants to get things done.

Responsibility: characterized by a person who likes to assume responsibilities, is very conscientious and worries a lot about what has to be done.

Self-sufficiency: characterized by a person who neither wants nor needs help, makes decisions by himself, is ashamed of accepting any help or favor.

Strict Moralism: characterized by a person who leads a very ethical life, is quite preoccupied by what is right and what is wrong, is inclined to follow well established rules.

Emotional Inhibition: characterized by a person who considers that to show emotion is a sign of weakness, is unable to express any form of emotion.

Conformity: characterized by a person whose interpersonal relationships are limited by emotional inhibition while communication takes place on a formal level. Acceptance of social demands, conventional attitude and no taste for change.

Many statements were constructed according to suggestions of psychologists who were asked to compose items related to the eight variables. Additional statements were derived from the Manifest Hostility Scale (Siegel, 1956) and from Murray's questionnaire (Murray, 1938). In the final form of the items, every attempt was made to reflect directly the behavioral observations and self-evaluations given in cases reported in the literature. To counter the effects of response-set, a number of items were phrased in the negative. The wording of the statements was not above the 5th-6th grade level of comprehension according to the Thorndike-Lorge Word Count (Thorndike & Lorge, 1944).

The content validity of the scale was established by the following method. Five clinical psychologists and five patients of the nonpsychosomatic population not included in the studied sample were asked to classify the 160 statements into eight categories according to the description of the variables. The agreement of 8 out of 10 judges was considered sufficient to determine the inclusion of a given item in the Index. The 148 statements selected by the above method were presented to 170 pilot Ss who were asked to rate themselves on a five-point scale for self and ideal self. Following Flanagan's technique (Flanagan, 1939), the final selection of the items was based on the correlation between item scores and total score for each variable. Items with a correlation less than .30 were eliminated. Fifteen items for each variable were finally retained.

Intervariable correlations were computed with the data collected for the item analysis. Correlations of .40 or greater were found between Responsibility and Efficiency (.59), Responsibility and Moralism (.45), Conformity and Moralism (.52), Aggression and Moralism (-.40).

The reliability coefficients, shown in Table 2, were established for each variable by correlating randomly selected halves and by applying the Spearman-Brown formula. The reliability of the Index was deemed sufficiently high to test the experimental hypotheses.

TABLE 2
COEFFICIENTS OF RELIABILITY FOR EACH VARIABLE
OF THE PEPTIC ULCER INDEX

Variable	Coefficient
Efficiency	.83
Aggression	.65
Moralism	.87
Emotional Inhibition	.80
Responsibility	.81
Dominance	.88
Conformity	.84
Self-sufficiency	.80

RESULTS

The general hypothesis that peptic ulcer patients, compared to a nonulcer population, endorse and value patterns of achievement and responsibility while denying patterns of dependency, was tested in terms of twelve specific variables. The data concerning each variable were treated with the *t* test in order to determine the significance of differences between experimental and control groups. The null hypothesis was rejected at the .025 level of confidence with a one-tailed test. Since the two control groups were found to be equivalent on all relevant variables, the test consisted of a comparison of the ulcer group with the combined control groups.

Self-ratings of the ulcer group, as measured by the Peptic Ulcer Index, were not statistically different from self-ratings of the non-ulcer samples for the following variables: Dominance, Aggression, Efficiency, Responsibility, Self-sufficiency, and Strict Moralism. However, the ulcer patients described themselves as more *emotionally inhibited* and more *conforming* than the control Ss at the .02 and .01 levels of confidence respectively, as shown in Table 3.

On the EPPS, no significant differences were found between experimental and control groups on the variables of Aggression, Dominance, Achievement, and Change. However the ulcer patients described themselves as less *intracceptive* than the control Ss as indicated in Table 4.

The second main hypothesis which stated that the discrepancy between self and ideal ratings is higher for the experimental group than for the combined control groups was not supported.

TABLE 3

MEANS, STANDARD DEVIATIONS, AND DIFFERENCES BETWEEN MEANS FOR EACH VARIABLE OF THE PEPTIC ULCER INDEX

Variables	Peptic Ulcer		Control 1		Control 2		Difference between Groups			
	Mean	SD	Mean	SD	Mean	SD	Peptic Ulcer Group vs. C ₁ & C ₂		C ₁ vs. C ₂	
							<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
Dominance	46.90	10.73	48.00	8.66	46.62	9.09	.19	<i>ns</i>	.51	<i>ns</i>
Aggression	40.80	8.94	43.15	12.08	39.51	8.04	.27	<i>ns</i>	1.38	<i>ns</i>
Efficiency	58.00	9.06	56.35	6.24	57.13	10.33	.66	<i>ns</i>	.30	<i>ns</i>
Responsibility	54.80	9.40	52.15	7.98	51.30	10.55	1.53	<i>ns</i>	.01	<i>ns</i>
Self-Sufficiency	56.03	7.79	52.00	8.03	52.93	9.91	1.94	<i>ns</i>	.36	<i>ns</i>
Moralism	58.00	8.69	53.70	11.36	55.75	11.70	1.49	<i>ns</i>	.65	<i>ns</i>
Emotional Inhibition	46.55	9.28	42.85	8.38	41.93	8.70	2.24	.02	.39	<i>ns</i>
Conformity	49.88	7.95	45.25	7.55	46.09	8.53	2.49	.01	.37	<i>ns</i>
Discrepancy Scores	99.97	38.70	109.85	45.07	102.20	45.45	.83	<i>ns</i>	.58	<i>ns</i>

Additional Finding

Among 10 other variables measured by the EPPS, no significant differences were found between experimental and control groups with the exception of Order. The ulcer patients described themselves as more orderly than control Ss ($p < .01$).

Additional Analysis of Data

An additional analysis of data was made to determine whether the results could be generalized to a civilian population. A comparison of the "career soldiers" (more than three years of service with intention to remain in service as a vocation) with the "noncareer soldiers" (draftees with no intention to remain in service beyond the obligated tour of duty) showed no significant differences. Therefore, the prob-

ability that the findings of the present study apply only to ulcer patients in the Armed Forces is quite limited.

DISCUSSION OF RESULTS

The data did not consistently support the defensive pattern of behavior proposed by Alexander. That is, peptic ulcer patients did not vigorously react against their longing for love and affection by expressing unusually strong needs for Achievement, Dominance, Aggression, Responsibility, Efficiency and Self-sufficiency. However, Emotional Inhibition, Avoidance of Insight and Conformity characterized the ulcer group. Such a differentiation suggests that ulcer patients may attempt to ward off the realization of their dependency needs by (a) an avoidance of in-

TABLE 4

t VALUES AND GROUP MEANS FOR EACH VARIABLE OF THE EPPS INVOLVED IN MAIN HYPOTHESIS

Variable	Peptic Ulcer		C ₁		C ₂		<i>t</i>	<i>t</i>
	Mean (<i>N</i> = 40)	SD	Mean (<i>N</i> = 20)	SD	Mean (<i>N</i> = 40)	SD	PU vs. C ₁ + C ₂	C ₁ vs. C ₂
Achievement	14.92	3.07	14.45	4.81	14.72	4.02	.39	.23
Change	14.42	4.72	14.70	6.06	15.25	4.40	.39	.40
Intracception	13.52	4.42	16.25	3.08	15.00	4.90	2.29*	1.04
Dominance	15.00	5.18	14.15	5.06	14.67	4.41	.57	.21
Aggression	14.42	4.28	14.55	4.67	12.65	3.18	.97	1.82

* Significant at .02 level of confidence.

sight into their motives and feelings, (b) an inhibition of expression of their emotional needs, and (c) an uncritical acceptance of social demands. This pattern can be seen as a compromise between an infantile need for protection and a socially acceptable, even sanctioned behavior. A conformist attitude is, in a way, a personally and socially acceptable form of dependency in which one can indulge without having to contend with disapproval and guilt.

The findings obtained might be accounted for in terms of a shift in behavior patterns socially acceptable at the present time in contrast with behaviors emphasized in the society of Alexander's patients. Uncritical acceptance of social demands, emotional inhibition, and avoidance of insight seem to constitute a meaningful and coherent cluster of personality variables. Stereotyped interpersonal relationships are characterized by a formal style of communication and group-oriented values. Several contemporary authors (Fromm, 1947; Riesman, 1950; Whyte, 1956) consider this need to conform and accept others' standards a major dictum in Western society at the present time. To characterize this pattern, Whyte has used the term "Social Ethic" while Riesman refers to "otherdirectedness." The dependency needs of the ulcer patient may be thus appeased, or partially gratified through the acceptable channel of conforming behavior. Under such circumstances, the Achievement drive and the subsequent conflict, Dependency-Achievement, are quite superfluous. However, 20 or more years ago, when Alexander formulated his theory, the socially predominant pattern which is reminiscent of Weber's Protestant Ethic (1930) included achievement, ambition, and competition. It is possible to speculate that the defensive pattern of behavior proposed by Alexander had only a temporary validity, and that the lasting meaning of this behavior pattern resided in the underlying motivation, the need to follow closely the culturally patterned set of values.

Several authors report similar findings concerning the conformity variable, without however attributing much significance to this particular aspect of the ulcer patient clinical portrait. Leary (1957) has noted the "hypernormal" behavior of ulcer patients. Harris,

Christiansen, and Ruesch (1946) mentioned conformity as a tentative form of adjustment. Weiner et al. (1957), in a multidisciplinary approach to the problem of peptic ulcer etiology, have found that, with the blood serum pepsinogen concentration controlled, the only criterion differentiating the ulcer cases was "the intensity of their attempt to maintain relationships with others" (p. 5). Thus, it may be the intensity of the need to relate to others, coupled with the inability to communicate effectively due to repressive mechanisms, which explains the emphasis on conformity in the personality of ulcer patients.

Although the unpredicted finding, strong order in ulcer cases, may be due to chance, it may be possible to account for its occurrence by considering that this variable is assessed in the EPPS as a methodical approach in work habits and in everyday living. Military life certainly enforces such an approach, and the high Order scores of the experimental group can be viewed as an additional nuance to the picture of a person with limited insight, little ability to express himself emotionally, and great need to conform.

In view of the fact that Alexander worked with patients of upper middle and lower upper class, business executives and intellectuals, and the fact that patients in the present study were lower middle class, skilled workers and semiprofessionals, it would seem profitable to extend this study to a larger socioeconomic group analogous to Alexander's sample. A study of this sort would help determine whether there has been a shift in the defense mechanisms of ulcer patients or whether the difference of findings is related to a difference in social class grouping.

SUMMARY

The defensive behavior pattern proposed by Alexander as characteristic of the ulcer patient in his handling of unconscious dependency needs was investigated. The existence of dependency needs was assumed, and the hypotheses tested concerned only the pattern of defenses against the realization of these dependency needs. It was hypothesized that ulcer patients would describe themselves as being more efficient, responsible, dominant, self-sufficient, achievement-oriented, aggress-

sive, moralistic, emotionality inhibited, conforming and resistant to change, but less in-traceptive and self-accepting. Forty ulcer patients were compared to 20 nongastrointestinal psychosomatic and 40 nonpsychosomatic patients, with the EPPS and a specially designed rating scale.

The results did not consistently support the defensive pattern proposed by Alexander and it can be said that the dependency conflict is no longer manifested in terms of dependency versus achievement. The ulcer patients are adopting behavioral patterns of frustration-avoidance in relation to their dependency needs rather than defending against these needs by a blatant pursuit of achievement and success. Dependency needs are defended against, yet partially satisfied through behavioral patterns which include personal and social constriction in the service of acquiescence to formalized social demands.

The results of the study were interpreted in terms of a shift in socially desirable patterns of behavior. Alexander's ulcer patients were viewed as inner-directed and governed by the Protestant Ethic while the ulcer patients of the present study were considered to be other-directed and dominated by the Social Ethic. The constant element in the ulcer patient's clinical portrait appeared to be his persistent need to follow closely the socially sanctioned set of values.

REFERENCES

- ALEXANDER, F. *Psychosomatic medicine*. New York: Norton, 1950.
- ALEXANDER, F., & FRENCH, T. M. *Studies in psychosomatic medicine*. New York: Ronald, 1948.
- BACON, CATHERINE. Typical personality trends and conflicts in cases of gastric disturbance. In F. Alexander & T. M. French (Eds.), *Studies in psychosomatic medicine*. New York: Ronald, 1948. Pp. 134-137.
- BENEDEK, THERESE. Control of transference relationship. In F. Alexander & T. M. French (Eds.), *Psychoanalytic therapy*. New York: Ronald, 1946. Pp. 173-206.
- BROWN, M., BRESNAHAN, T. J., CHALKE, F. C. R., PETERS, D., POSER, E. G., & TOUGAS, R. V. Personality factors in duodenal ulcer: A Rorschach study. *Psychosom. Med.*, 1950, 12, 1-5.
- FLANAGAN, J. C. General considerations in the selection of test items and a short method of estimating the product-moment coefficient from data at the tails of the distribution. *J. educ. Psychol.*, 1939, 30, 674-680.
- FROMM, E. *Man for himself*. New York: Rinehart, 1947.
- GARMA, A. On the pathogenesis of peptic ulcer. *Int. J. of Psychoanal.*, 1950, 31, 53-72.
- HAMBURGER, M. A revised occupational scale for rating socio-economic class. Unpublished manuscript, Columbia Univer., 1957. (Mimeo.)
- HARRIS, R. E., CHRISTIANSEN, C., & RUESCH, J. Character and conflict in Naval enlisted personnel with duodenal ulcer. *Amer. Psychologist*, 1946, 1, 263.
- KAPP, F. T., ROSENBAUM, M., & ROMANO, J. Psychological factors in men with peptic ulcers. *Amer. J. Psychiat.*, 1947, 103, 700-704.
- LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.
- MODELL, A. H., & POTTER, H. Human figure drawings of patients with arterial hypertension, peptic ulcer and bronchial asthma. *Psychosom. Med.*, 1949, 11, 282-292.
- MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
- PRINCE, S. D. A comparative study of personality characteristics in bronchial asthma and peptic ulcer patients as revealed by the Rorschach test. Unpublished doctoral dissertation, Univer. of Southern California, 1950.
- RIESMAN, D. *The lonely crowd*. New Haven: Yale Univer. Press, 1950.
- SIEGEL, S. M. The relationships of hostility to authoritarianism. *J. abnorm. soc. Psychol.*, 1956, 52, 368-372.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia Univer., 1944.
- UNITED STATES DEPARTMENT OF THE ARMY. Administering and scoring the Army Classification Battery. *USA Pamphlet*, 1948, No. 611-100.
- VAN DER HEIDE, C. A study of mechanisms in two cases of peptic ulcer. In F. Alexander & T. M. French (Eds.), *Studies in psychosomatic medicine*. New York: Ronald, 1948. Pp. 206-223.
- WEBER, M. *The protestant ethic and the spirit of capitalism*. London: George Allen Unwin, Ltd., 1930.
- WEINBERG, H. An investigation of unconscious conflict in cases of peptic ulcer. Unpublished doctoral dissertation, Harvard Univer., 1951.
- WEINER, H., THALER, MARGARET, REISER, M. F., & MIRSKY, I. A. Etiology of duodenal ulcer. *Psychosom. Med.*, 1957, 19, 1-10.
- WHYTE, W. H. *The organization man*. New York: Simon & Schuster, 1956.

(Received June 5, 1959)

INHIBITION AND PERCEPTION OF MOVEMENT ON THE RORSCHACH

ANN FILINGER NEEL

Wyandotte County Guidance Clinic, Kansas City, Kansas

Validation of hypotheses underlying Rorschach variables is an elusive task. Yet, despite the problems of definition, and experimental control, research aimed at clarifying basic issues is being conducted. A series of recent studies by Meltzoff, Singer, Korchin, and their associates are a case in point. Generalizing from Werner's (1945a; 1945b) sensory-tonic theory of perception and Rorschach's (1942) speculations about movement these investigators predicted a positive relationship between motor inhibition and perception of human movement on the inkblots. They produced inhibition variously by tying subjects down (Korchin, Meltzoff, & Singer, 1951), and by having them write slowly, in a constricted fashion (Meltzoff, Singer, & Korchin, 1953). Their hypothesis was supported. Other types of movement revealed more ambiguous associations with inhibition. Further studies (Singer & Herman, 1954; Singer, Meltzoff, & Goldman, 1952; Singer & Spohn, 1954; Singer & Sugarman, 1955) produced similar results and suggested that both the capacity to inhibit and the ability to perceive movement might be related to healthy childhood experiences and "good" identification with parents.

The present study is an effort to test the generality of these findings. In the first place, the earlier studies were confined to motor behavior, but the authors seemed to imply their results could pertain to all varieties of inhibition including suppression of ideas and impulses. This assumption was never tested. Secondly, the ambiguous relationship between inhibition and animal or inanimate movement requires clarification. Finally, the original studies of Meltzoff, Singer, and Korchin utilized the responses to only two Rorschach cards, one under normal conditions counterbalanced with another under experimental

conditions. While later investigations employed the entire Rorschach protocol, the association between perceived movement and inhibition was never tested on all 10 cards. To substantiate the relationship it should be placed in the perspective of the total Rorschach situation. This becomes especially important in view of an apparent tendency to increase the number of responses to the second card viewed, regardless of conditions.

These considerations gave rise to the hypotheses of this study:

Hypothesis 1

Any situation calling for inhibition, or producing an inhibited state, should increase the amount of human movement perceived on the Rorschach.

Hypothesis 2

Any situation calling for inhibition, or producing an inhibited state, should increase the amount of animal movement perceived on the Rorschach.

Hypothesis 3

Any situation calling for inhibition, producing an inhibited state, should increase the amount of inanimate movement perceived on the Rorschach.

EXPERIMENTAL DESIGN

Subjects

Ninety-five introductory psychology students were involved in the study. Two subjects failed to complete the experiment, making the final $N = 93$.

Procedure

The use of all 10 Rorschach cards made it impossible to follow the practice of making each person his own control. Instead, subjects were randomly assigned to the several treatments comprising the ex-

periment. To control for the influence of intelligence on perception of human movement, the groups were matched on Wonderlic (1945) scores. There was no attempt to control age and sex.

In all, five different conditions were employed. Two of these were similar to the inhibiting conditions of the Meltzoff, Singer, Korchin studies. The Tied group, $N = 18$, had their dominant arm bound to their desks during the Rorschach. The Slow Writing group, $N = 19$, were instructed to copy as slowly as possible a printed paragraph describing a foot race. Ten minutes was allotted for the task. The paragraph was substituted for the short phrase used by the earlier investigators in hopes of increasing the inhibitory effects. All other groups copied this material but were given no instructions as to speed of writing.

The third experimental condition was introduced to test for the effects of inhibition of ideas. To produce such a state, a Behavior Inhibition group, $N = 19$, received the following supplement to the usual Rorschach instructions: "Normal people commonly see at least a few sexual things on each picture, private parts of people's bodies or various kinds of sexual activity. Despite sexual ideas being so common in this experiment you must not give me anything pertaining to sex no matter how clearly you see it."

Before one can assume that such instructions produce inhibition, one must discover whether the suggestion regarding the normality of sexual percepts increases the probability of their occurrence. For this reason, a Sex Control group, $N = 18$, were given the set that sex responses were commonly seen, and were not told to avoid them.

A Normal Control group, $N = 19$, was given the ordinary introduction to the Rorschach.

All five conditions were run concurrently in one large classroom, this form of data collection not only being efficient, but providing a class demonstration of multivariate design. All instructions were written and bound with necessary materials in booklet form. The subjects were told in the beginning that the study involved a number of different tasks and that the people next to them would not be doing the same things they were. They were requested to ignore the activities of their neighbors with the promise that the study would be explained later. Each Rorschach card was shown on a projection screen for two minutes. The subjects were asked to give two responses per card. No inquiry was conducted.

The Rorschach protocols were scored blind, guided by the criteria Klopfer (Klopfer, B., Ainsworth, Klopfer, W. G., & Holt, 1954) has established except that human percepts were not given a score of M unless movement was explicitly described. Form level was scored according to the list given by Beck (1949, pp. 159-195). Because there was no inquiry some error is undoubtedly present in assigning form level scores. It was believed, however, that the information produced by such scores was sufficiently useful to justify an attempt at applying it.

Additional determinants were included in the final tabulation of scores, but were counted as half a score. This process of weighting multi-determinant responses, and the fact that many subjects did not give twenty responses resulted in unequal R s. To equate individual records, frequency counts for each score were converted into percentages of total number of responses.

RESULTS AND DISCUSSION

With certain exceptions dictated by procedure followed in the earlier studies, the chi square statistic was used. Fourfold tables were constructed, the median score of the Control Group being used as the cutting point for the comparison of it with the various other groups. Persons scoring on the median were included in the below median cell. Yate's correction was used whenever expected frequency in any cell was less than 10.

Prior to the consideration of the results it is necessary to discover if certain contaminating factors could influence the findings.

Controls for Extraneous Variables

Intelligence. In addition to matching for Wonderlic scores, each group was checked to discover if there was an indication of any relationship between good human movement responses ($M+$), and IQ, or between all movement responses ($M+$ and $M-$) and IQ. Equal numbers of persons fell in all four cells of the chi square tables, revealing no relationship between human movement responses and intelligence in the population of the present experiment.

Number of Responses. Differences in the number of responses or determinants could affect the number of movement responses especially in the Meltzoff, Singer, Korchin (1953) approach. In that study the number of responses given under inhibitory conditions by the whole group was compared to the number of movement responses given under normal conditions. No significant difference in R was found in the earlier studies, although a slight increase was noted on the second card seen.

In the present study, the number of responses ranged from 9 to 20 in the Behavior Inhibition group and the Normal Control group, from 10 to 20 in the other groups. The median number of responses for the en-

tire sample was 18. There was no significant difference in the number of responses given by each group, but there was a consistent trend toward more responses in experimental groups. Only 32% of the Normal Control group gave more than 18 responses, compared to 37% of the Tied group, 39% of the Sex Control group, 45% of the Slow Writing group, and 47% of the Behavior Inhibition group.

There was no significant difference in the number of determinants used, but the same consistent trend was in evidence.

Validity of Inhibitory Procedures. We cannot discover whether the various experimental conditions employed did indeed produce inhibitions but we can determine whether the instructions had any effect. The Behavior Inhibition group were given a suggestion to see sexual percepts but were asked to avoid giving them. The Sex Control group were given the suggestion to see such percepts. The Normal controls were given no suggestions at all. If the various instructions had any effect, the Behavior Inhibition group should produce less sex responses than the Normal Control group. Both groups should give fewer sex percepts than the Sex Control group. Chi square comparisons yield the predicted relationships, significant beyond the 5% level for a one tail test where $df = 1$ (χ^2 Normal Control group vs. Sex Inhibition group = 2.92; χ^2 Normal Control group vs. Sex Control group = 7.27; χ^2 Sex Inhibition group vs. Sex Control group = 17.13).

The Tied and Slow Writing Inhibition groups should not be affected since they received normal Rorschach instructions. Chi square comparisons of the number of sex responses given by these groups with the Normal Control group were not significant as predicted. On the basis of these statistical comparisons it appears that instructions regarding normality of sexual percepts tended to produce such percepts. Also, it appears that those subjects who were told to avoid sexual material did so and therefore inhibited their activity insofar as we can tell from this indirect evidence.

In like manner we can test the effects of the slow writing instructions. The group instructed to write slowly should copy signifi-

cantly fewer words than any other group. Chi squares were all significant in the predicted direction at beyond the 2% level for a one tail test where $df = 1$ (χ^2 Slow Writing group vs. Tied group = 4.20; χ^2 Slow Writing group vs. Sex Inhibition group = 8.19; χ^2 Slow Writing group vs. Sex Control group = 5.94; Slow Writing group vs. Normal Control group = 4.20).

Human Movement

The preliminary checks disposed of, we can now turn to the test of the major hypotheses. The first hypothesis held that human movement responses should increase with inhibition. During the tabulation it became obvious that it might be well to separate $M+$ from $M-$ and test for each one. The first three rows of Table 1 summarize the results of these comparisons.

Only the results of the Normal Control vs. Sex Inhibition group on $M+$ are significant, but the other tests suggest tendencies in the same direction. Contrary to the hypothesis and the previous findings good movement decreased with inhibition. The findings were dissimilar for $M-$. There was an insignificant trend toward increase in $M-$ in the Sex Inhibition and Sex Control groups. If such a trend were "real," it seems most likely due to instructions about sex responses. Calculations for combined $M+$ and $M-$ are not significant but again showed the consistent decreasing tendency.

The results so far definitely do not fit with the previous findings but this may be because of different statistical tests. The Meltzoff, Singer, Korchin study (1953) used a t test of the difference between the number of movement responses before and after inhibition. Since the number of responses and the number of determinants did not differ significantly we can perform a similar test although the tendency to increase in responses and determinants casts some doubt on the validity of this procedure. The means and variances for the groups are given in Table 2. The F tests for homogeneity of variance showed significant differences between the Control group and all others except the Slow Writing group. Under these circumstances the t test could not be used and the Mann Whitney U test

TABLE 1
CHI SQUARE COMPARISONS OF EXPERIMENTAL GROUPS WITH CONTROL GROUP:
PERCENTAGE OF VARIOUS MOVEMENT RESPONSES GIVEN

	Control Group	Slow Writing Group ^a	Tied Group ^a	Sex Inhibition Group ^a	Sex Control Group ^a	All Experimental Group ^b
<i>M</i> + Median Percentage	25.5	19	20.5	15	19.5	
χ^2 with Control Group		N.S.	N.S.	6.27	N.S.	8.33
<i>P</i> Value				<.02		(<.10).05
<i>M</i> - Median Percentage	0	0	0	4.5	5.67	
χ^2 with Control Group		N.S.	N.S.	N.S.	2.18	N.S.
<i>P</i> Value					(<.10).05	
Total						
<i>M</i> Median Percentage	26.67	22	22.5	23	24	
χ^2 with Control Group		2.92	2.71	N.S.	N.S.	N.S.
<i>P</i> Value		(<.10).05	.10			
<i>FM</i> + Median Percentage	5	6	10	8	7	
χ^2 with Control Group		N.S.	4.50	1.73	N.S.	6.37
<i>P</i> Value			<.02*	<.10*		<.10*
<i>FM</i> - Median Percentage	1	5	4	2	2	
χ^2 with Control Group		1.69	4.66	N.S.	N.S.	N.S.
<i>P</i> Value		(<.10)*.05	<.02*			
Total						
<i>FM</i> Median Percentage	12	13	16.67	13	12.5	
χ^2 with Control Group		N.S.	2.37	N.S.	N.S.	N.S.
<i>P</i> Value			(<.10)*.05			
<i>m</i> Median Percentage	1	6	4	7.2	5.5	
χ^2 with Control Group		2.63	6.32	11.30	3.30	13.61
<i>P</i> Value		(<.10)*.05	<.01*	<.001*	(<.05)*.02	<.01*
Total <i>M</i> , <i>FM</i> & <i>m</i>						
Median Percentage	42	42	42	42	39.5	
χ^2 with Control Group		N.S.	N.S.	N.S.	N.S.	N.S.
<i>P</i> Value						

^a *df* = 1.^b *df* = 4.

* One tail test.

was substituted. Neither *t* nor *U* gave significant results for any of the groups. The only finding of importance was the increased variance in the experimental groups. It would seem that inhibiting conditions may make Rorschach subjects behave less consistently.

Meltzoff, Singer, and Korchin also made a *t* test of the total amount of movement pro-

duced, and such a test was likewise made with the data from this study. The means and variances for distribution of all movement responses are given in Table 2. Again, no significant results were obtained but there was a consistent trend toward more movement in the experimental groups. This might be a manifestation of the phenomenon observed by

TABLE 2

MEANS, VARIANCES, AND TESTS OF HOMOGENEITY OF THE DISTRIBUTIONS OF *M* RESPONSES AND TOTAL MOVEMENT PRODUCTION

	Control Group	Slow Writing Group	Tied Group	Sex Inhibition Group	Sex Control Group
<i>M</i> Mean	4.00	3.45	4.67	3.66	4.20
<i>M</i> Variance	1.80	1.82	6.35	5.46	6.19
<i>F</i> test for Homogeneity of Variance with Control Group		1.00	3.52	3.03	3.44
<i>P</i> Value		N.S.	<.02	<.02	<.02
<i>M</i> , <i>FM</i> & <i>m</i> Mean	6.50	7.21	9.11	7.92	7.44
<i>M</i> , <i>FM</i> & <i>m</i> Variance	7.24	7.03	15.11	10.43	13.46
<i>F</i> test for Homogeneity of Variance with Control Group		1.03	2.09	1.44	1.86
<i>P</i> Value		N.S.	N.S.	N.S.	N.S.

the earlier investigators, dulled to insignificance by alterations and experimental procedure. On the other hand it might be a reflection of the tendency to increase the number of responses. After all, this combined movement score contained nearly half of the responses in the individual protocol according to the last row of Table 1. The results of the test *FM* and *m* will offer additional reasons for these increases.

In an effort to duplicate the Meltzoff, Singer, Korchin findings, the Slow Writing group was tested for relationship between the amount of inhibition shown and the number of *M* responses. A fourfold chi square table was constructed comparing the status of a subject on the movement distribution with his position in the inhibition distribution (number of words written). The results of the comparison were not significant.

Since the Meltzoff, Singer, Korchin studies used only two Rorschach cards (III and VII or VIII and IX), the number of movement responses were tabulated for each card. Inspection of the totals for each card revealed

differences of only one or two points. Use of the entire Rorschach apparently did not camouflage the contribution of any individual card.

Animate and Inanimate Movement

The hypotheses regarding *FM* and *m* also predicted increases. Animal movement was scored for good and poor form. Tests were made utilizing these scores both separately and combined. Rows four through seven of Table 1 summarize these statistical tests. The tests for *FM*+ and *FM*- were far from universally significant, but all pointed toward increases in these determinants under conditions calling for inhibition. The combination of *FM*+ and *FM*- reflected the increase but none of the chi square tests using the combined score were significant. While the comparisons involving animal movement were hardly an imposing demonstration of the predicted relationship, they at least revealed consistent trends which occasionally reached significance.

The results for *m* were unequivocally posi-

tive, even for the Sex Control group where inhibition, if produced, was an experimental artifact.

The chi square test for all types of movement showed that there is no apparent difference among groups. The contradictory trends demonstrated among the various types of movement could easily have cancelled each other out, thus rendering the use of this particular combination of movement of questionable value.

DISCUSSION

The hypothesis regarding human movement emerged from this study with little to support it. The most likely conclusion is that human movement responses are not affected in any consistent manner by the experimental procedures used here. It is possible that good movement responses may actually decrease while poor responses show some slight tendency to increase when the possibility of sexual material is increased. Variability among persons must be expected. Providing that differences in technique are not too great for generalization and that the present results are not chance variations, it is possible that the increases observed by Meltzoff, Singer, and Korchin could be due in part to increased responsiveness or to the effects of animal and inanimate movement.

The results of the comparisons for *FM* and *m*, as a whole, support the predictions about them. The findings in regard to animal movement were not significant in many cases, but demonstrated trends in the direction hypothesized.

It will be recalled that Klopfer holds animal movement to be indicative of awareness of impulses in which the person does not necessarily indulge. By itself, however, it does not necessarily imply inhibition. Klopfer interprets animal movement to mean frustration only in the absence of *CF* and without an optimal relationship to *M* (Klopfer et al., 1954, p. 266). The inconsistent findings may result from the interference of these other variables. The possibility of an interrelationship with color responses was demonstrated by Singer and Spohn (1954) in regard to *M*.

Klopfer interprets *m* as indicative of tension and conflict due to efforts at inhibiting

the need and integrating it into behavior. The present results support such a prediction. The fact that the Sex Control group also gave increased *m* is not necessarily inconsistent, for nothing about the experimental design assures that tension and conflict are decreased by the suggestion to perceive sexual content.

SUMMARY

On the basis of previous studies and extrapolation therefrom, it was hypothesized that any situation calling for inhibition should increase the amount of human movement, animal movement, and inanimate movement on Rorschach protocols. To test these hypotheses, five groups of college students were used. Total *N* = 93. In one experimental group inhibition was induced by requesting the subjects to copy a paragraph in long hand as slowly as they could. In another, inhibition was induced by binding the subject's dominant arm to his chair. A Behavior Inhibition group was instructed that it was normal to see sexual things on the test but was asked to avoid giving such percepts. Another group was employed as a control to see if the suggestion that sexual responses were normal had any effect on protocol. A fifth group was a control group for whom all conditions were normal. The predicted increases in human movement percepts were not observed. In fact there was an insignificant tendency toward a decrease of good human movement. The findings in regard to animal movement were in the predicted direction, but were not ordinarily significant. The results of inanimate movement were unequivocally positive, validating both the present hypothesis and Klopfer's interpretation of *m* responses. The failure of predictions may have been due to changes in experimental procedure or to a general increase in responsiveness and variability under experimental conditions.

REFERENCES

- BECK, S. J. *Rorschach's: Test I Basic Processes*. New York: Grune & Stratton, 1949.
- KLOPPER, BRUNO, AINSWORTH, MARY D., KLOPPER, WALTER G., & HOLT, ROBERT R. *Developments in the Rorschach Technique*. Vol. I. *Technique and Theory*. New York: World Book, 1954.

- KORCHIN, S. J., MELTZOFF, JULIAN, & SINGER, J. L. Motor inhibition and Rorschach movement responses. Paper read at American Psychological Association, Chicago, September, 1951.
- MELTZOFF, JULIAN, SINGER, JEROME L., & KORCHIN, SHELDON J. Motor inhibition and Rorschach movement responses: a test of the sensory-tonic theory. *J. Pers.*, 1953, 21, 400-410.
- RORSCHACH, H. *Psychodiagnostics*. New York: Grune & Stratton, 1942.
- SINGER, JEROME L., & HERMAN, JACK. Motor and fantasy correlates of Rorschach human movement responses. *J. consult. Psychol.*, 1954, 18, 325-331.
- SINGER, J. L., MELTZOFF, J., & GOLDMAN, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.*, 1952, 16, 359-364.
- SINGER, JEROME L., & SPOHN, H. E. Some behavior correlates of Rorschach's experience type. *J. consult. Psychol.*, 1954, 18, 1-9.
- SINGER, JEROME L., & SUGARMAN, DANIEL A. A note on some projected familial attitudes associated with Rorschach movement responses. *J. consult. Psychol.*, 1955, 19, 117-119.
- WERNER, H. Motion and motion perception: A study in vicarious functioning. *J. Psychol.*, 1945, 19, 317-327. (a)
- WERNER, H. Perceptual behavior of brain-injured mentally deficient children. *Genet. Psychol. Monogr.*, 1945, 31, 51-110. (b)
- WONDERLIC, E. F. *Wonderlic Personnel Test Manual*. Northfield, Illinois, 1945.

(Received June 8, 1959)

REJECTION OF FALSE INFORMATION ABOUT ONESELF AS AN INDICATION OF EGO IDENTITY

WALTER GRUEN¹

Beth Israel Hospital, Boston, Massachusetts

In 1950 Erikson described personality development as a series of stepwise changes in the ego from birth to death. Each of his eight stages of development was defined as a crisis point which the maturing ego could solve in one of two polar directions, so that one of the solutions would always lead to a widening of ego boundaries and hence towards further growth.

Since then several studies have been made or been started to either verify the theory or to use Erikson's concepts in an exploration of individual differences alongside other systematic dimensions. Gruen's² and Yufit's (1956) studies are examples of both of these. A further attempt was made in the present study to demonstrate a connection between behavior predicted from Erikson's theory and the particular personality integration one had reached during the ego identity stage of development.

Erikson (1950) defines ego identity as follows:

The growing and developing youths, faced with this physiological revolution within them, are now primarily concerned with attempts at consolidating their social roles. They are sometimes morbidly, often curiously, preoccupied with what they appear to be in the eyes of others as compared with what they feel they are and with the question of how to connect the earlier cultivated roles and skills with the ideal prototypes of the day. . . . The integration now taking place in the form of ego identity is more than the sum of the childhood identifications. It is the inner capital accrued from all those experiences of

each successive stage, when successful identifications led to a successful alignment of the individual's basic drives with his endowment and his opportunities. . . . The sense of ego identity, then, is the accrued confidence that one's ability to maintain inner sameness and continuity (one's ego in the psychological sense) is matched by the sameness and continuity of one's meaning for others (p. 216).

We followed Erikson's assumption that ego identity once achieved results in a more stable personality organization which is visible in behavior. Hence the hypothesis was advanced that a person with ego identity will show this accrued stability of a now integrated role pattern by rejecting evaluations of himself by others, if these evaluations do not coincide with his own crystallized notions about himself. Similarly, a person still casting about for his various identities and roles and values, who shows signs of ego diffusion, will be more prone to accept others' evaluation of himself, even if they may have no demonstrable relationships with any of his fleeting self-images. In fact, he is prone to grasp them like a drowning man reaching for a straw, especially if such evaluations come from status sources.

In an effort to obtain a more objective measure of ego identity than the ratings based on interviews and projective data first employed by Gruen (see Fn. 2) and Yufit (1956), it was decided to utilize the real-ideal *Q* sort discrepancy correlation score developed by Butler and Haigh (1954). Support from this choice of measure comes from Erikson's definition of ego identity as the acceptance of a stable set of roles which then become the ego ideals of the individual. The acquisition of standards and their full acceptance into the ego would then reduce discrepancies between idealized attributes and attributes seen as characteristic of oneself.

¹ Formerly at University of Buffalo.

² There is under preparation by the research staff of the Committee of Human Development of the University of Chicago a book on the adult person in urban life in which appears a chapter by W. Gruen: *The Investigation of Adult Personality Through an Experimental Application of Erikson's Theory of Ego Development*.

Further indications for the rationale of this measure here come from Turner and Vanderlippe's (1958) comparison between real-ideal correlations and various indices of adjustment including sociometric ratings strongly suggestive of Erikson's ego identity syndrome. A tentative validation was also attempted here by constructing 14 questionnaire items around the theme of ego identity versus ego diffusion. A total ego identity score on the 10 items on which 45 subjects (Ss) distributed themselves evenly, correlated -0.42 with real-ideal discrepancy correlations on the *Q* sort.

The specific hypothesis tested in this study was as follows: A high self-ideal discrepancy as an indication of low ego identity would be related to an uncritical acceptance of a fake personality sketch which presumably was a realistic appraisal for the *S*.

METHOD AND RESULTS

Forty-five summer session students 18-24 years old were asked to sort the 100 *Q* sort items in the usual manner, using first real self and then ideal self instructions. When they reassembled after one week, they were given the above mentioned questionnaire and were then handed individually typed personality sketches. They were told that these sketches were tentative conclusions from the testing of the previous week. The fake sketches were all identical in content but not in form to prevent discovery of the former while glancing at one's neighbor. The sketch consisted of 13 statements which Forer (1949) had derived from astrology magazines and had used as a pedagogic device in the classroom.

To allow for rejection of the sketch, the Ss were told that the original test was a relatively new and untried one and that we wanted an opinion on how well our conclusions coincided with their own ideas as a way of finding out more about the test. On an attached sheet they were asked to rate on 10-point scales (a) how effective our test had been in revealing personality and (b) how revealing our sketch had been in pointing out important characteristics of their personality. After the sheets were collected, the ruse was pointed out and the experiment was thoroughly explained.

The correlation scores between the two *Q* sortings resulted in a range of scores from $+0.87$ to -0.43 , with a mean of $+0.43$ and a sigma of $.29$. Looking at the spread of ratings on the two 10-point scales derived from Forer (1949) two thirds of the Ss were distributed among the upper four points or "acceptance" poles. The correlation between the "effective" and the "revealing" ratings was so high (0.80) that only the "revealing" score was used in the final analysis because it also was the one score appropriate to the hypothesis.

To test the study hypothesis the correlation between the sortings by each *S* was correlated with his "reveal" reaction to the fake personality sketch. Since the spread of ratings on the latter measure occurred within a small range of points around a mean of 8.0 a chi square was also computed between the independent and dependent variable by splitting each distribution into two halves at the median.

The correlation between the correlation for the two *Q* sorts and the degree of acceptance of the fake sketch was -0.45 . A correlation of -0.39 for this *N* occurs less than once in a 100 by chance. The corollary chi square resulted in a value of 3.92 , significant at the 0.05 level. Therefore, the hypothesis about the behavioral visibility of high and low ego identity is confirmed, if ego identity is indeed a useful concept and if it is adequately measured by a real-ideal discrepancy in attitudes about oneself.

The correlation between the "reveal" ratings and the score based on the aforementioned ego identity questionnaire items was $-.26$. This figure is not quite significant at the $.05$ level, but is in the expected direction.

DISCUSSION

The findings are first of all a verification of one part of Erikson's theory, insofar as the demonstrated connections between the discrepancy of real-ideal self and the willingness to accept any assertions about oneself were derived from the theory. The results also have implications for a better understanding of the uncertainty behavior of adolescents and young adults. It would not be too farfetched to speculate that fad changes, vocational win-

dow shopping and the indiscriminant parade of dating partners are all related to willingness to accept glib statements about oneself and hence manifestations of ego diffusion, just as the opposites might be indications of ego identity. If self-ideal discrepancy is indeed a good and valid measure of this dimension, we might expect it to correlate with objective measures of these behavior patterns typically found in many adolescents.

One disappointing note is the lack of a significant relationship between degree of acceptance of the fake sketch and the a priori questionnaire items which were devised to measure ego identity. Either the questionnaire medium is not a good one to catch a relatively complex dimension like ego identity, or both of these two measures have so much variance specific to them, that they in turn correlate with the self-ideal discrepancy criterion, but not with each other. A factor-analytic investigation of an expanded question-

naire for Erikson's eight ego stages on several 100 Ss in all age ranges is now in process for a future report to answer some of these questions.

REFERENCES

- BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 55-75.
- ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.
- FORER, B. R. The fallacy of personal validation: A classroom demonstration of gullibility. *J. abnorm. soc. Psychol.*, 1949, 44, 118-123.
- TURNER, R. H., & VANDERLIPPE, R. H. Self-ideal congruence as an index of adjustment. *J. abnorm. soc. Psychol.*, 1958, 47, 202-206.
- YUFIT, R. I. Intimacy and isolation: Some behavioral and dynamic correlates. Unpublished doctoral dissertation, Univer. of Chicago, 1956.

(Received June 8, 1959)

DIAGNOSTIC PROTOTYPES AND DIAGNOSTIC PROCESSES OF CLINICAL PSYCHOLOGISTS¹

MANDEL RUBIN²

Cleveland State Hospital

AND FRANKLIN C. SHONTZ

Highland View Hospital, Cleveland, Ohio

The present study is based upon the premise that more should be known about the process of clinical judgment and psychological evaluation. Asch's original studies (1946) are to be credited with much of whatever current interest exists in the general problem of how people form conceptions of each other. Shontz (1956), has stressed elsewhere the need for focalizing this interest upon the specific diagnostic processes of the clinical psychologist.

Some of the basic questions which require empirical investigation are:

1. To what extent may psychologists be expected to agree on certain basic psychodiagnostic conceptions?
2. To what extent is the diagnostic process influenced by diagnostic prototypes or preconceptions?
3. Is there a common diagnostic process which is consistent regardless of the individual psychologist, or of the individual patient in question?
4. If such a common diagnostic process exists, to what extent is it influenced by the kind of information upon which it is based and by the order in which this information is obtained?
5. To what extent may psychologists working independently with identical data be ex-

pected to agree in their final diagnostic conclusions?

The present authors see the answers to these questions as being dependent upon at least three classes of variables: (a) those related to the characteristics of the patient being evaluated, (b) those related to the characteristics of the psychologists who perform the evaluation, and (c) those relating to the type of patient data which is made available to the psychologist, and the order of their presentation to him.

The present study was limited to examining the effects of variables of the third class, holding relatively constant the relevant patient and psychologist factors. The method imposed heavy restrictions upon the generalizability of results in terms of variables of the first two types; but it had the distinct advantage of permitting the investigators to get an extremely "close look" at the complex evaluative process in a well-defined, if highly specific, research situation. Patient variables were controlled by limiting the subjects to male paranoid schizophrenics hospitalized in a single institution. Psychologist variables were controlled through the use of a highly homogeneous group all of whom had considerable formal and informal contact with each other, and most of whom had received their education at the same university. Only the types of patient data and the orders of presentation of these data were experimentally varied.

METHOD

Measurement

A Q sort instrument was made up of 54 items selected by two psychologist judges from a list of

¹ The senior author is indebted to D. W. Miles, W. M. Taylor, M. Horowitz, and C. Waltner of Western Reserve University for advice and guidance in completing the dissertation upon which this research is based. Thanks are also due to Cleveland State Hospital for the use of its facilities.

² The senior author is now at the Department of Marriage Conciliation, Court of Common Pleas, Cuyahoga County, Cleveland, Ohio, and the junior author at the University of Kansas.

120 items constructed by Stephenson to reflect Rorschach features describing various aspects of personality (Beck, 1954). Items were selected to be most consonant with the clinical picture of paranoid schizophrenia.³ In describing a patient or a diagnostic concept, each psychologist rater was required to sort the final 54 items into a quasinnormal distribution with the following frequency categories: 1, 2, 4, 6, 9, 6, 4, 2, 1. Usual methods were employed to convert these data into statistically appropriate quantitative form (Stephenson, 1955).

Raters

The raters were 24 psychologists, 16 of whom were regular staff members of psychology departments of either state or county hospitals. Three of the 16 were department heads. Three other raters worked for institutions offering primarily counseling or advisory services. Five were either students or interns, and 1 was in private practice. Eight raters were female, 17 male. Eleven were PhDs in psychology and 11 others had completed virtually all requirements for the doctorate except for the dissertation; the remaining 2 had completed at least two years of appropriate graduate work. The age range of the raters was 23 to 58, with an average of 33.5 years. Their clinical experience averaged slightly over 6 years, with a range of 1 to 19 years. All but one of the raters had received or were receiving their graduate education at the same university.

At one point, the examiner served as an additional rater of all patients. At the time of the research he had completed three years of graduate study, had five years' experience, and was 38 years of age.

Patients

Twenty-five Ss were selected from the inpatient population of a local state hospital. All patients had been assigned the diagnosis of *paranoid schizophrenia*, and the group consisted entirely of males. Age range of this sample was 28 to 62 years (median: 41 years); and length of time in the hospital ranged from 5 to 76 months (median: 38 months). All but seven patients had been in mental hospitals one or more times previously.

One of the patients served as a special *control patient*. The patient was 41 years of age, had been hospitalized 63 months, with no previous hospitalizations, and had 10 years of education.

Procedures

a. Each rater was first asked to use the *Q* sort instrument to describe his conception of the prototypic male paranoid (*Pd*), catatonic (*Ct*), and hebephrenic (*Hb*) schizophrenic. Homogeneity of these

prototypic descriptions was evaluated by correlational methods, and composite *Q* sort descriptions of the three relevant diagnostic categories were established from the ratings. Reliability of the instrument was partially established by requiring eight of the raters to repeat their descriptions of the paranoid schizophrenic.

b. All raters (except *E*) were exposed to case study information on the control patient. Information was presented in four increments, and raters made *Q* sort descriptions of the patient after each increment. The increments, in order of presentation, were:

I. Statistical data: patient's code number, birthdate, date of hospital admission, education, occupation, marital status, religion, race, sex, source of admission, number of previous hospitalizations.

II. Interview: tape recording of a thirty-minute, semistructured interview with the patient. All interviews were conducted by *E* to control for skill and style of interviewing.

III. Psychological report and test protocols: a psychological report based on a minimum battery consisting of the Rorschach and the Wechsler-Bellevue (Form I) Vocabulary, Block Design, and Similarities subtests. In most cases this minimum battery was supplemented by various other tests, selected by the examining psychologist. Testing and report writing were done by regular members of the hospital staff in their ordinary course of duty.

IV. Social case history: a full report made by the psychiatric social worker on the basis of information obtained from an informant, usually a close relative of the patient.

Increments were presented in the order listed above, and this order of presentation was considered standard, since it represents the usual process of clinical examination in the institution. Data on actual diagnostic conclusions were not made available to the raters.

Q sort descriptions obtained from all raters on the control *S* were treated statistically to investigate changes from each step to the following steps of the procedure.

Q sort descriptions were also correlated with the composite prototypic diagnostic concepts obtained in Step *a* to determine the relative influence of these concepts at varying stages in the diagnostic process.

c. Since Step *b* does not enable one to distinguish whether changes in conception are a function of the type of information or of the serial order in which the information is presented, each rater (except *E*) was next assigned a separate subject from the sample of patients. The procedure of Step *b* was repeated, except that information increments were presented in orders which differed systematically from rater to rater.

Intercorrelations between successive descriptions provided data for analysis of variance procedures which permitted the evaluation of serial effects and information sources as contributors to the final *Q*

³ A table giving the complete *Q* sort has been deposited with the American Documentation Institute. Order Document No. 6286, remitting \$1.25 for 35 mm. microfilm or \$1.25 for 6 by 8 in. photocopies.

TABLE 1

Q SORT ITEMS MOST AND LEAST CHARACTERISTICS OF THE PROTOTYPIC DIAGNOSTIC CONCEPTIONS

Diagnostic Prototype	Most Characteristic Items	Least Characteristic Items
Paranoid Schizophrenic	9 Projects (ascribes his own unpalatable or painful wishes or fears to persons or objects in his world) 13 Delusional thoughts, systematized 17 Characterized by intellectualization	14 Inattentive to point of distractibility 2 Thought processes are accidental, tangential 3 Hallucinatory, with much ego disorganization
Catatonic Schizophrenic	37 Blocks out external perceptions and reacts to stimuli from within 46 Characterized by stereotypy in action 50 Affect is blocked (<i>i.e.</i> as a defense)	21 Breadth of interest spontaneity is increased (<i>ie.</i> what might be called undirected spontaneity) 13 Delusional thoughts, systematized 17 Characterized by intellectualization
Hebephrenic Schizophrenic	3 Hallucinatory, with much ego disorganization 2 Thought processes are accidental, tangential 6 Much intrusion of accidental material	17 Characterized by intellectualization. 18 Hallucinatory, but with little ego disorganization (<i>i.e.</i> low level regressed ego disorganization) 54 Overconventional

sort descriptions. *E* also reviewed all available data on each patient and provided *Q* sort descriptions of each, based upon his full knowledge of the case material. Correlation of these descriptions with those provided by the other raters provided further reliability data.

RESULTS

The homogeneity values (average interjudge *r*'s for the composite diagnostic prototypes were .47 for *Pd*, .40 for *Ct*, and .54 for *Hb* (all values significant at the .01 level).

The intercorrelations of the composite prototypes derived from these descriptions were as follows: *Pd* vs. *Ct*, $-.27$ (not statistically significant); *Pd* vs. *Hb*, $-.54$ (significant at $P < .01$); *Ct* vs. *Hb*, 0.00 . Some of the items considered by the raters to be most and least descriptive of each composite prototype are presented in Table 1.

Sort vs. re-sort reliabilities for the eight raters who repeated their descriptions of the *Pd* prototype ranged from .74 to .92, with a mean of .86.

Median intercorrelations between successive *Q* sort descriptions of the control patient are presented in Table 2. Nonparametric analysis of variance, using methods suggested by Wilcoxon (1949) showed that these correlations could be separated into three statistically homogeneous subgroups, as indicated

by the code letters assigned to each coefficient in Table 2.

Correlations of the diagnostic prototypes with *Q* sort descriptions of the control patient provided after the presentation of each increment of case study information are presented in Table 3.

Table 4 presents the median correlation coefficients between successive *Q* sorts of the 24 paranoid schizophrenic patients, according to the type of case study information used and the serial position in which the information was presented. Nonparametric analysis

TABLE 2
MEDIAN CORRELATION BETWEEN SUCCESSIVE
Q SORT DESCRIPTIONS OF THE
CONTROL PATIENT

Serial Positions Compared	Median Correlation	Code
I vs. II	.58*	A
I vs. III	.54	A
I vs. IV	.50	A
II vs. III	.70	B
II vs. IV	.65	B
III vs. IV	.77	C

* All values significant at .01 level.
Note.—Code letters indicate coefficients which fall into statistically homogeneous groups.

TABLE 3
AVERAGE INTERCORRELATIONS BETWEEN
COMPOSITE DIAGNOSTIC PROTOTYPES
AND *Q* SORT DESCRIPTIONS OF
CONTROL PATIENT

Serial Position	Composite Paranoid Prototype	Composite Catatonic Prototype	Composite Hebephrenic Prototype
I	.80**	.00	-.26
II	.65**	.07	-.28*
III	.70**	-.15	-.21
IV	.81**	-.21	-.22

* Significant at .05 level.

** Significant at .01 level.

Note.—All values are means, derived by way of *r* to *z* transformation.

TABLE 4

MEDIAN CORRELATIONS BETWEEN *Q* SORT DESCRIPTIONS AS A FUNCTION OF TYPE OF CASE STUDY INFORMATION AND SERIAL POSITION IN WHICH INFORMATION WAS PRESENTED

Type of Information	Serial Position Intercorrelations		
	I × IV	II × IV	III × IV
Statistical data	.46*	.70	.82
Interview	.72	.82	.91
Psychological report	.70	.81	.86
Case history	.62	.67	.82

* All values significant at .01 level.

Note.—Final description (Serial Position IV) taken as basis for comparison of serial effects.

of variance of the raw data upon which Table 4 was based showed significant differences to exist between serial positions (*p* less than .01) and nearly significant differences to exist between types of case study information. The interaction was not statistically significant.

Correlations between the raters' final descriptions of their patients and *E*'s descriptions of these same patients ranged from a low of .28 to a high of .79. All but the lowest correlation were significant at the .01 level. The lowest correlation was found between *E* and the single judge who was not trained at the same university as the rest of the raters.

DISCUSSION

The present study showed that the group of psychologists examined possessed signifi-

cantly homogeneous prototypes of the three diagnostic categories, Paranoid, Catatonic, and Hebephrenic Schizophrenia. Homogeneity values were highest for *Hb* prototype, indicating least controversy with respect to the characteristics of this group of patients. Since hebephrenics are often considered to be the most deteriorated patients, this finding lends support to Shontz's observation (1954, pp. 83-87) that the more obviously abnormal a patient is, the easier psychologists find it to agree upon a descriptive diagnosis. It is also possible that the relatively lower homogeneity value found for the *Ct* prototype was due to the concentration of some raters upon phenomena associated with catatonic stupor, while other raters may have been more concerned with the excited states that often characterize this type of patient. It may be that differences in homogeneity of diagnostic prototypes are related to the "coherence" of the diagnostic category itself. Thus, for example, one might also expect less homogeneity to exist for the category "manic depressive psychosis" than for the category "depression" alone.

The rather moderate homogeneity values for all the diagnostic prototypes leaves considerable room for disagreement between raters on specific behavioral or dynamic items. It suggests the possibility that for each prototype there are probably a few items which are consistently placed in specific *Q* sort categories, but that many other items exist which may be placed almost anywhere without destroying the basic integrity of the diagnostic picture. Future studies might well be directed toward the isolation and specification of such crucial items for the various diagnostic categories.

It is obvious from the intercorrelations between composite diagnostic prototypes that the catatonic prototype is independent of the other diagnoses. The negative correlation between *Pd* and *Hb*, however, suggests that these are thought of by this group of psychologists as rather opposite diagnoses. Examination of the items in Table 1 reveals the differences to lie in the emphasis upon intellectual rigidity in the paranoid and upon intellectual disorganization in the hebephrenic.

It is interesting to note in Table 3 how these diagnostic prototypes correlated with *Q* sort descriptions of the control patient. It is quite apparent that all raters identified the patient as a paranoid schizophrenic even on the basis of the statistical data alone (in Serial Position I). Doubtless this was deduced from the raters' knowledge of the sheer frequency of occurrence of this diagnosis in state hospitals, although it is also possible that the statistical data itself suggested specific things that led to this hypothesis. Of equal interest is the observation that the paranoid prototype tended to exert relatively less influence on the conceptualizations after the interview was heard, and then regained its high correlation with the *Q* sort descriptions as more case study material was added. It is striking that the final descriptions correlated almost exactly as high with the diagnostic prototype as did the initial descriptions, based on relatively little information. There is the strong suggestion in these data that the diagnostic prototype did, in fact, serve as an hypothesis which was established early in the procedure and which tended generally to become more firmly entrenched as confirming case study material became available. The slight, though steady, trend away from the *Ct* prototype, which is also evident in these data, is also of interest and suggests that a definite process of differential judgment was operative in the psychologists' descriptions. It seems to imply, in short, that the reactivation of the paranoid prototype in Serial Position IV was not the result of sheer stereotypy of the psychologists' thinking but that this reactivation did arise from a process of considered discrimination.

Table 2 shows a little more clearly how this process took place. Each *Q* sort description was found to correlate significantly with all other descriptions. However, the correlations between Description I and the following descriptions are far lower than the correlations between Description II and the descriptions following it, and so on. Thus it may be said that, while there was a basic continuity of concept formation throughout the diagnostic process, definite changes of conceptions were also taking place. These changes were greatest following Serial Position I, they were

relatively less following Serial Position II, and they were least following Serial Position III. It is therefore obvious that the psychologists were being influenced by the information presented them. It is also obvious, however, that they were not sufficiently influenced to alter significantly their basic conception of the control patient as a paranoid schizophrenic individual.

Three questions arise about these conclusions, and all three are answered in Table 4. The first question is whether these results might not be a function of the individual patient selected. Since the data in Table 4 are based upon *Q* sort descriptions obtained on 24 patients, and since the serial effects are the same (that is, for each type of case study data, the I vs. IV correlation is always less than the II vs. IV correlation, which in turn is always less than the III vs. IV correlation, yet all values remain significant at the .01 level), it is clear that the individual patient was not the determining variable. A conception of the patient was consistently formed early in the diagnostic process and this conception tended to remain holistically the same. More detailed information altered the conception only in its more specific aspects. The second question is whether the trends observed are a function of the order in which case study information was presented to the raters. Since the results of Table 4 are based upon all possible orders of presentation of the data, and since the trends remain the same, the answer to this question is obviously that it is not a function of the order of presentation of information. The third question is whether the phenomenon, which has been called the serial position effect, is dependent in any way upon the type of case study information used. Although the interaction in Table 4 was not found to be statistically significant, there was a very definite tendency for general correlation values to be lowest when statistical data were presented first and highest when the interview was presented first. That is, the initial conceptions of the patient underwent more change over-all in the former situation, but the serial position effect was constant nevertheless. Apparently, the most stable initial conceptions were induced by immediate ex-

posure to a recording of the patient himself, a fact which suggests that written material, no matter how detailed, is not a complete substitute for actual contact with the patient, where psychological diagnosis is concerned.

One final question remains unanswered, i.e., to what extent the final patient descriptions provided by the raters are unique functions of the raters themselves. The generally significant correlations between *E*'s *Q* sort descriptions of the 24 *Ss* and the descriptions provided by the raters is convincing evidence that interjudge reliability of these conceptions is generally satisfactory. It is impossible to tell from the present study whether the single low interrater reliability which was found is actually a function of differences in education as such, or whether it represents another, perhaps, deeper type of personal difference between *E* and the other psychologist in question. A number of factors could be responsible, among which could be (a) the effects of possible emotional involvement of the psychologist with the patient, and (b) the effects of the rater's general mode of approach to the data available. (For example, it is possible that the presentation of information, inconsistent with an initial diagnostic conception, may induce one psychologist to alter his judgments completely, while another psychologist might prefer to reject the contradictory data and retain earlier judgments.) As previously mentioned, the restrictions of the present study with respect to many important patient and psychologist parameters are recognized. Despite these limitations, it is fair to say that the findings suggest several basic principles which can, and should, be tested further on more widely varying groups of *Ss* and professional raters.

SUMMARY

A *Q* sort investigation was made of clinical psychologists' conceptions of three specific diagnostic categories (paranoid, catatonic, and hebephrenic schizophrenia) and of the process of concept formation which is involved in arriving at descriptive diagnosis of individual patients. Twenty-five psychologists served as raters and 25 male paranoid schizophrenics served as *Ss*. Case study data were presented to the psychologists in such a way

as to make possible the evaluation of the effects of four types of clinical information and the effects of serial position, or order, in which this information was presented.

It was found that the psychologists employed possessed significantly homogeneous prototypic diagnostic conceptions, and they tended to view paranoid and hebephrenic schizophrenia as somewhat opposite disease entities; catatonic schizophrenia appeared as a statistically independent conception.

A consistent and significant serial position effect was found in the diagnostic process of actual patients. It was apparent that psychologists formed a fairly complete clinical picture of the patient quite early in the diagnostic process and that this basic conception changed in its details but not in its basic configuration as more information was provided to the raters. The most stable initial clinical conceptions were formed when recorded interviews with the patient were presented first.

Evidence was also found that individual differences in psychologists may, in certain instances, have profound effects upon the reliability of the diagnostic process. It is not yet certain what type of individual differences are most important in this regard, although it is possible that the psychologist's education, his emotional involvement with the patient, or differences in the degree to which he is willing to be influenced by contradictory information, play a part in the process. Further research on these phenomena is strongly recommended.

REFERENCES

- ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258-290.
- BECK, S. J. The six schizophrenias: *Reaction patterns in children and adults*. Research Monograph No. 6. New York: Amer. Orthopsychiat. Ass., 1954.
- SHONTZ, F. C. Clinical psychologists' conceptualizations of their patients. Unpublished doctoral dissertation, Western Reserve Univer., 1954.
- SHONTZ, F. C. Evaluative conceptualizations as the basis for clinical judgments. *J. consult. Psychol.*, 1956, **20**, 223-226.
- STEPHENSON, W. *The study of behavior*. Chicago: Univer. Chicago Press: 1955.
- WILCOXEN, F. Some rapid approximate statistical procedures. New York: Amer. Cyanamid Co., 1949.

(Received June 9, 1959)

WECHSLER-BELLEVUE PERFORMANCES OF REMITTED AND UNREMITTED ALCOHOLICS, AND THEIR NORMAL CONTROLS

F. PLUMEAU, S. MACHOVER, AND F. PUZZO

Alcoholic Clinic, Downstate Medical Center, State University of New York

The study presented here is part of a larger research project investigating personality variables involved in remission from alcoholism. Because of obvious differences between remitted and unremitted alcoholics simply on a functioning level, as well as in their responses to similar opportunities for recovery and their ability to adapt to a new role, it was hypothesized that some part of a remitted group's recuperative characteristics might be reflected in their intellectual performances on a standard intelligence test.

Prior studies (Halpern, 1946; Kaldegg, 1956; Peters, 1956) have indicated that alcoholic populations which were available for study generally earned IQs higher than the general adult population. In addition, these studies report few total test or subtest deviations from normal population intellectual parameters. However, there is little information available on subgroups in alcoholism. This paper bears on the aforementioned studies and presents data from an additional population of "alcohol-freed" subjects (Ss).

PROCEDURE

The Wechsler-Bellevue Intelligence Scale for Adults (Form I) was administered to a group of 22 male chronic alcoholics in remission (active AA members with a minimum of 2 years of sobriety); a group of 23 male unremitted alcoholics (never more than six months of sobriety); and a group of 23 male normal controls (all no more than social drinkers). Administration of the tests was by the same examiner to all Ss. The three groups were equated for age, ethnic background and education. For the remitted group the age range was 27 to 55 years; for the unremitted, 28-56; and for the controls, 28 to 54. The average age for each group was 42 years. Thirteen of each of the experimental groups were of Irish extraction and 15 of the normal controls were of similar background. The remainder

were distributed over a diversity of ethnic backgrounds, including Anglo-Saxon, German, and Eastern European. The average highest school grade reached was 12 for the remitted alcoholics, 11 for the unremitted group and 11 for the normal controls. The remitted group's education ranged from 8 to 20 school years; the unremitted 8 to 18; and the controls 8 to 18. The two experimental groups were, in addition, equated for duration and intensity of alcoholism, and duration of exposure to Alcoholics Anonymous. The average duration of alcoholism in the remitted group was 19.27 years and 19.39 years in the unremitted group. The duration of alcoholism ranged from 6 to 29 years in the remitted group and from 6 to 38 years in the unremitted group. The average number of years contact with AA was 5.39 for the remitted group and 5.26 for the unremitted Ss. Neither of the experimental groups, at the time of testing, contained members who were currently hospitalized or in clinic attendance.

The weighted subtest means, and Verbal, Performance and Total IQ means were subjected to analyses of variance. In addition, each S in each group was assigned a scatter index score on the basis of 11 subtests. The scatter index is computed by calculating the mean weighted score of the 11 subtests. The total deviation, disregarding sign from this mean is the index. The Rank-Sums test (Senders, 1958) was used to check the significance of the differences found between the groups on this variable.

RESULTS AND DISCUSSION

The mean subtest scores and mean Verbal, Performance, and Total Scale IQs for each of the three groups are given in Table 1. The range of Total Scale IQs for the remitted group is 99-139, for the unremitted group, 105-134, and for the normal controls, 99-139. The mean Total Scale IQ for each group is above the average for the general adult population. This finding for alcoholics is consistent with the findings of previous studies.

The analysis of variance of each set of mean values revealed only one significant F ratio and that on the Arithmetic subtest, at the .01 level of confidence. The subsequent t tests between the three groups on this variable revealed significant differences between the remitted and unremitted alcoholics and between the unremitted alcoholics and the normal controls, both at the .01 level. In both cases the unremitted group had the lower mean.

Thus, our two alcoholic populations reveal but one significant difference on the Wechsler-Bellevue variables investigated. The higher performance of the remitted as compared with the unremitted alcoholics in Arithmetic may reflect a more effective concentration consistent with the better stabilized sobriety of the remitted group and with their more disciplined adaptation to the 12-step program of Alcoholics Anonymous. It should be noted, however, that the running of 14 analyses of variance in such a homogeneous set of subsamples increases the chance finding of significant differences and therefore decreases confidence in this subtest difference.

TABLE 1

WECHSLER-BELLEVUE SUBTEST AND TOTAL IQ
MEANS FOR THE THREE GROUPS

Subtest	Remitted ($N=22$)	Unremitted ($N=23$)	Controls ($N=23$)
Information	13.81	13.27	13.39
Comprehension	13.50	13.14	12.83
Digit Span	10.04	11.14	11.61
Arithmetic	13.36	10.77	13.43
Similarities	13.00	12.09	12.13
Vocabulary	14.32	13.95	13.87
Picture Arrangement	9.59	9.73	10.70
Block Design	11.77	10.95	11.96
Object Assembly	10.68	11.32	12.26
Digit Symbol	9.81	9.45	10.48
Verbal IQ	121.45	117.82	121.09
Performance IQ	117.77	116.14	120.00
Total IQ	121.27	118.45	121.83

TABLE 2

RANK-SUMS TEST RESULTS COMPARING THE
THREE GROUPS ON THE SCATTER INDEX Z 's

	Remitted	Controls
Remitted Unremitted	.42	3.31* 2.93*

* Significant at .01 level.

The results of the scatter index analysis are shown in Table 2. The data do not show any differences between the remitted and unremitted groups, but each experimental group differs from the normal controls. The alcoholic groups produced significantly higher scatter index scores than the normal group ($P = .01$). The fact that the two alcoholic groups are identical in mean scatter and that each independently differs significantly from the normal controls is evidence for the reliability of this finding.

The diagnostic implications of this scatter measure as reflective of severe psychopathology was noted in a recent study by Trehub & Scherer (1958). Using the same scatter index as employed in this study, these authors successfully differentiated schizophrenics from neurotics and character disorders. Using a cutting score of 19 for schizophrenia, they report 70% successful diagnosis. Our combined alcoholic groups contain 33 members out of 45 with scatter indices of 19 or above, a substantial 73%. Eight or 33% of our normal controls have scatter indices that high. This, of course, does not imply that our alcoholic groups are therefore largely schizophrenic. Since our alcoholic groups were screened for presence of psychosis or brain pathology by means of extensive conferences and examination by psychological and psychiatric staffs it seems likely that the results of the scatter index are not related to severe psychopathology or clinically detectable brain damage. The possibility, however, that prolonged alcoholic excess, in terms of effecting subtle brain damage, might have contributed to this finding is a moot point. A partial check on this was attempted by running a Pearson r between the number of years each alcoholic reported he had been drinking in

excess and his scatter index score. The correlation obtained was .05; a nonsignificant finding. In addition, no differences between the controls and alcoholics were found on those subtests usually associated with organicity, i.e., Block Design, Object Assembly, Digit Span, etc. An investigation of the determinants of the alcoholics' Wechsler-Bellevue scatter, and a similar investigation of the scatter of a more pathological group (schizophrenics for example) would provide a better basis for understanding the nature of contributing factors to these uneven intellectual profiles.

Thus, the independence of subtest scatter from duration and stability of remission, and perhaps also from years of prolonged drinking, suggests that unevenness of intellectual pattern might be due to deeply rooted personality characteristics of the chronic alcoholic. It may represent a long-term process of erratic intellectual development and functioning symptomatic of a more generalized deviation in personality organization and in the patterns of adaptation, which antedate the addiction to alcohol.

SUMMARY

Male remitted and unremitted alcoholics and their normal controls, equated for age, education and ethnic background, were administered the Wechsler-Bellevue Intelligence Scale for Adults (Form I).

The total IQs and weighted subtest means of both alcoholic groups generally supported previous studies.

Significant differences in only one subtest were found. On the Arithmetic subtest, the unremitted alcoholics scored lower than both the remitted alcoholics and normal controls ($P = .01$).

A scatter index analysis revealed no significant differences between the remitted and unremitted alcoholics. Both experimental groups, however, had significantly higher scatter index scores than the normal controls ($P = .01$). These differences suggest a pattern of intellectual unevenness in the alcoholic deviant from that of a normal population.

The significance of these findings was discussed and possible explanations were offered.

REFERENCES

- HALPERN, FLORENCE. Studies of compulsive drinkers: Psychological test results. *Quart. J. Stud. Alcohol*, 1946, 6, 468-479.
- KALDEGG, A. Psychological observations in a group of alcoholic patients. *Quart. J. Stud. Alcohol*, 1956, 17, 608-628.
- PETERS, G. A. Emotional and intellectual concomitants of advanced chronic alcoholism. *J. consult. Psychol.*, 1956, 20, 390.
- SENDERS, VIRGINIA L. *Measurements and statistics*. New York: Oxford Univer. Press, 1958.
- TREHUB, A., & SCHERER, I. W. Wechsler-Bellevue scatter as an index of schizophrenia. *J. consult. Psychol.*, 1958, 22, 147-149.

(Received June 12, 1959)

A RATIONALE FOR SELECTING TOYS IN PLAY THERAPY

HAIM G. GINOTT

Child Guidance Clinic, Jacksonville, Florida

Child therapists of different orientations make use of toys and play materials in the diagnosis and treatment of emotionally disturbed children. The rationale for this practice is the belief that play is the child's symbolic language of self-expression. Through the manipulation of toys the child can state more adequately than in words how he feels about himself and the significant people and events of his life. To a considerable extent, the child's play is his talk and the toys his words.

If this assumption is true, then the kind of toys and materials found in the playroom is an important therapeutic variable. Yet there are very few guideposts in the literature to the proper equipping of therapeutic settings. In many clinics, the playroom looks like a junkyard with an astonishing assortment of toys of various vintages, some of them marred and disfigured beyond use. New toys are amassed at random, with the therapist's personal predilection being the decisive factor in their acquisition. The following "confession" will serve as an illustration: "I suffer temptations in toy stores . . . and I have no difficulty in persuading myself that a certain doll is just what I need for a certain four-year-old who has been getting along just fine without it" (Fraiberg, 1954, p. 3).

The literature on toy selection is contradictory, and with few exceptions (Lebo, 1955, 1958) is based on intuition rather than investigation. For example, Arthur (1952) and Slavson (1952) advocate a playroom with many enticing toys, whereas Fraiberg (1954) thinks that therapy proceeds more effectively when the child is given only a few toys and is not absorbed in fancy materials and fascinating activities. There is no agreement about the value of specific materials, nor even about the necessity of having a special room set aside and furnished for child therapy.

Clearly, a consistent rationale is needed for the selection of materials that contribute

maximally to the process of child treatment. The present paper aims to formulate criteria for selecting effective toys and tools for play therapy with children under the age of 10.

THERAPEUTIC OBJECTS AND OBJECTIVES

The aim of child therapy, as of all therapy, is to effect basic changes in the intrapsychic equilibrium of the child. Through relationship, catharsis, insight, reality testing, and sublimation, therapy brings a new balance in the structure of the personality, with a strengthened ego, modified superego, and improved self-image. The value of any toy, object, or activity in child therapy depends on its contribution to the realization of these objectives. In evaluating activities and materials, one should consider their effect on the inner processes of therapy. There are five major criteria for selecting and rejecting materials for child therapy. A treatment toy should: (a) facilitate the establishment of contact with the child; (b) evoke and encourage catharsis; (c) aid in developing insight; (d) furnish opportunities for reality testing; (e) provide media for sublimation.

Tools for the Therapeutic Relationship

A therapeutic relationship can be established and maintained only if the therapist understands the child's communications. The assumption is that everything said or done by the child in the playroom has meaning and significance in his frame of reference. Yet it is not easy even for experienced therapists to understand all of the child's play messages. Questioning the child about the meaning of his play is worse than futile; it results in resistance and silence. Appropriate toys make it easier for the therapist to understand the meaning of the child's play. Thus, for example, children usually play out family scenes by using dolls that represent mother, father, and siblings. In the absence of such

dolls, a child may symbolically play out family themes by using big and little wooden blocks. But the exact meaning of the message may escape the therapist. Banging two blocks together may represent spanking, or intercourse, or may merely be a test of the therapist's tolerance for noise. Inserting a pencil into a pencil sharpener may represent intercourse, but it may also mean that the pencil needs sharpening. However, when a father doll is put on top of a mother doll, the therapist has less room for misinterpretation. For the child, pencil and doll may be equally useful as a means of expression, but to the therapist they are not. The presence of a doll family enables the child to assist the therapist in understanding him. Some materials are superior for use in therapy because they give therapists tools for understanding the child, without diminishing the fluidity of his play or conversation. It seems reasonable to assume that a playroom with a variety of selected toys is more conducive to therapeutic communication than one filled with odds and ends of junk.

Many therapists find it easier to structure therapy and to make contact with a child when the playroom contains materials whose very presence reflects permissiveness. Some toys and tools are associated in children's experience with parental and sibling restrictions and punishments. Many children have been scolded or spanked for messing with mother's typewriter, fooling with brother's flashlight, or playing with father's tool kit. Nothing conveys permissiveness to these children as much as the presence of such materials in the playroom for their own use. Noise-making toys such as drums, pegboards, xylophones, air rifles, and cap guns serve the same purpose. They communicate loud and clear the adult's basic spirit of tolerance.

If it is at all possible, every child should find in the playroom some toy, tool, material, or activity refused to him in the past. Perhaps the case histories of clinic children should contain more detailed information on the child's play deprivations and aspirations.

Toys for Catharsis

The common assumption that children project their emotional needs on any play materials is only half true. It ignores the fact that

playroom materials have behavior-propelling qualities of their own. Some toys elicit the expression of children's needs and problems, whereas others limit them. By a wise choice of toys, the therapist augments his power to catalyze and control the play sessions. Materials determine directly the choice of activity and indirectly a connected chain of potential events. An activity once started leads to a more or less predictable sequence of behavior. Thus, beyond finger painting lurk events that are likely to occur because of the nature of the materials and the activity. Finger painting is likely to stimulate children: in the course of painting they will get some paint on themselves. They will want to clean up then and there. In the process, they will splash water on themselves, the floor, and the therapist. Feelings will be heightened; there will be complaints, apologies, cautions.

Such a chain of events largely determines the particular character of the therapeutic relationship. It determines whether the therapist intervenes out of choice or necessity, and whether he must provide liberty or limits. The anticipated sequence of actions and reactions can be utilized to increase the specificity of treatment plans for children with different needs and symptoms. Since the nature of social interaction is drastically affected by variation in activities and materials, these variables can be manipulated to regulate the degree of mildness or wildness of the playroom. Specific decisions will depend on each child's nuclear needs and on the therapeutic intent, which is to help children achieve a healthy balance between acting out and self-restraint.

Another misconception which is frequent among child therapists is the belief that all acting out has therapeutic value. It is true that catharsis in children almost always involves motility and acting out. Yet it cannot be overemphasized that acting out per se has no curative effects beyond pleasure and release. Acting out in children does not usually lead to self-evaluation, recognition of motivation, guilt, and attempts to alter behavior. This is true for all persons with inadequate superegos, but it is especially evident in young children, in whom a weak superego and strong narcissism make acting out just fun. This is one reason why acting out by

young children has to be limited through therapeutic intervention (Ginott, 1959). Acting out is of value only when it represents working out of the child's core difficulties. Knowing each child's central problems and providing relevant toys and tools is an important, though often a neglected, part of a therapist's responsibility.

In planning for therapeutic catharsis the therapist should furnish materials that elicit acting out related to the child's fundamental problems and should avoid materials that evoke diffuse hyperactivity. Thus hyperkinetic, overactive, and brain damaged children should not find finger paints or running water in the playroom. Such materials overstimulate them and invite smearing of each other and the walls. And although such behavior may be highly pleasurable for children, it is psychonoxious in effect and leads to weakened inner controls and disorganization of personality. These children need materials and activities that will focus rather than diffuse their flow of energy. Pounding pegboards, building block houses, shooting rifles, driving nails, sawing wood, or constructing boxes may give form and direction to the disorganized urges of these children. These activities expose the children to sustaining interests that call for concentration and coordination. In an atmosphere of wise encouragement the children's frustration will be enhanced and they will become able to focus energies on projects and goals, both in and out of the playroom.

Fearful, fragile children should find in the playroom materials that can be handled without the aid of tools, such as water, paint, sand, play-dough, dolls, chalk, crayons, etc. They need materials that will enable them to conceal what they do not want to reveal, and to do and undo acts without detection and embarrassment. Thus, they can use clay for sculpturing or scalping, sand for building or burying, and paint for decorating or dirtying. The reversibility of these media enables children to change the identity of their symbolic revelations at will and makes it safe for them to explore their inner and outer worlds.

Materials for Insight

Toys do not contribute directly to the achievement of insight. This is only gained

through growth in inner security. But toys do facilitate interpretations that help children become more aware of themselves and their relationships with significant people. It is through this awareness that insight is attained.

There will be occasions on which it is desirable to arrange the materials in the playroom so as to provoke a child's habitual mode of reaction. In general, one would not aimlessly repeat within the playroom the frustrations and deprivations that the child is encountering outside. But occasionally frustrating situations may be set up in order to enable children to gain insight into the dynamics of their behavior. When only two attractive guns are provided for three aggressive children, the probability is that conflict will occur. The children's heightened reactions in the conflict situation offer the therapist a chance to confront each child with his habitual, unfruitful, self-defeating modes of reaction and to suggest more effective ways of handling conflict.

Toys for Reality Testing

From the moment he enters treatment the child tests the realities of the setting, the reaction of the therapist, the response of his group-mates, the nature of the materials. The frustrations and satisfactions encountered in handling playroom materials and the sense of power acquired in mastering them have direct bearing on the child's ego strength and self-image. The ego is strengthened and the self-image is enhanced by realistic success and both are damaged by repetitive failure. While young children need encounters with reality as part of their treatment, it must be so arranged as to suit the child's changing powers. Accordingly, it is especially important that children who come to treatment with vast experience in defeat do not relive failure in the playroom. Too many playrooms contain complicated mechanical toys that are hard to wind, guns that are difficult to cock and puzzles that are impossible to put together. Such play materials produce frustration and dependence, and hinder the development of feelings of mastery and security. A playroom should contain materials of graded difficulty which allow each child, regardless of handicap, to achieve some measure of suc-

cess. Frequently a child with a long history of failure will discover that he had more ability in handling materials than he has ever known. The consequent improvement in self-image may mark the beginning of "cure."

Materials and activities can contribute to the development of a realistic social outlook. Axline (1955) asks the pertinent question: "If play-therapy is an experience in self-exploration, . . . how does the child learn to expand beyond his self-centeredness to a recognition and appreciation of others?" (p. 622). In order for play therapy to be an experience in social learning, children should be provided with situations and materials that demand exploration of others as well as themselves. Most children, at certain times in their therapy, should be exposed to peers, resistive materials, and planned scarcity of tools so that they can test themselves in relation to social actualities. A setting that selectively and intentionally provides one electric jigsaw for three children who work with wood is likely to create frustration and conflicts and subsequent modes of conflict-resolution. The therapeutic gains in such a setting will be greater than those where materials are provided indiscriminately.

Media for Sublimation

One of the important objectives of psychotherapy is to help children to develop sublimations that are compatible with society's demands and mores. Our culture does not give children much choice about relinquishing infantile gratifications. Children are forced to give up interest in primary processes as early as possible. Children are strictly forbidden to handle body products and even the desire to explore them is prohibited. This task is usually accomplished by punishment, which leads to repression. It is of the utmost importance in therapy to offer children opportunities to enjoy forbidden pleasures in acceptable substitute ways. Since "a small child's imagination makes urine of every fluid and feces of every messy substance" (Balint, 1954, p. 24), sand, water, paint, and clay provide excellent means of sublimating urethral and anal drives. No playroom for small children is complete or adequate without such materials. Enuretic children should be given

paint and running water, encopretic children, mud and brown clay. Firesetters should have capguns, sparklers, and flashlights. All children should find in the playroom miniature utensils for cooking and serving meals to sublimate oral needs, dolls that can be dressed and undressed to sublimate sexual needs, and punching bags, targets and guns to sublimate aggressive drives. Each child should have the opportunity to express his needs *symbolically* in a great variety of ways according to his changing capacities. A child should be led to express anger by punching dolls and destroying clay figures, as well as by composing critical poems and writing murder mysteries. The lack of adequate outlets in the playroom may impede the emergence of more mature means of sublimation. The therapy setting must provide materials that allow growth in the repertory of self-expression.

SUMMARY

This article presents a consistent rationale for selecting toys that contribute maximally to the process of play therapy. A desirable treatment toy: (a) facilitates the establishment of contact with the child, (b) encourages catharsis, (c) aids in developing insight, (d) furnishes opportunities for reality testing, and (e) provides media for sublimation.

REFERENCES

- ARTHUR, HELEN. A comparison of the techniques employed in psychotherapy and psychoanalysis of children. *Amer. J. Orthopsychiat.*, 1952, 22, 484-498.
- AXLINE, VIRGINIA M. Play therapy procedures and results. *Amer. J. Orthopsychiat.*, 1955, 25, 618-626.
- BALINT, ALICE. *The early years of life*. New York: Basic Books, 1954.
- FRAIBERG, SELMA H. *Psychoanalytic principles in casework with children*. New York: Family Service Ass. of America, 1954. (Pamphlet)
- GINOTT, H. G. The theory and practice of "therapeutic intervention" in child treatment. *J. consult. Psychol.*, 1959, 23, 160-166.
- LEBO, D. The expressive value of toys recommended for nondirective play therapy. *J. clin. Psychol.*, 1955, 11, 144-148.
- LEBO, D. A formula for selecting toys for nondirective therapy. *J. genet. Psychol.*, 1958, 92, 23-34.
- SLAVSON, S. R. *Child psychotherapy*. New York: Columbia Univ. Press, 1950.

(Received June 15, 1959)

CHANGE IN ATTITUDES AND DEGREE OF VERBAL PARTICIPATION IN GROUP THERAPY WITH ADULT OFFENDERS¹

ALEXANDER B. SMITH,

Kings County Court, Brooklyn, New York

ALEXANDER BASSIN,

BARO Clinic, Brooklyn, New York

AND

ABRAHAM FROEHLICH

Hofstra College

There has been some speculation concerning the relationship between degree or amount of verbal production and therapeutic change during group or individual therapy. Psychotherapy has often been called "the talking cure" and the assumption that there is a distinct positive relation between the amount of the patient's verbalization and the extent of improvement has appeared so obvious a truism as to hardly require explicit statement. However, there is little in the literature bearing on this matter. Gibby, Stot-sky, Hiler, and Miller (1954) have shown that the more verbally productive patients remain longer in psychotherapy, thereby presumably having greater opportunity for improvement. On the other hand, Rogers (1942) suggested that therapy may actually take place during periods of silence. Powder-maker and Frank (1953) cautioned that in group therapy "in itself speech may have no more value than silence."

The present study was an investigation of the relationship between degree of verbal participation and changes in attitudes in a therapy group of adult offenders on probation. It was hypothesized that individuals

with a greater degree of verbal participation would evidence greater positive changes in their attitude toward authority figures and social conformity.

PROCEDURE

Fifteen consecutive probation placements whose ages ranged from 16 through 32 years, whose IQs ranged from 80 to 125, and whose educational level ranged from the fifth to the twelfth grade, were divided into two groups of eight and seven subjects (Ss). Each group met for 90 minutes weekly for 15 sessions at the Brooklyn Association for the Rehabilitation of Offenders Clinic, a noncourt clinic, for client-centered group therapy. An observer kept track of the amount of time each individual spoke during each session, using a record blank which divided each 90-minute session into 20-second intervals.

In order to evaluate the changes in attitudes, a modification of the TAT was employed following Stern (1952) and Newman (1953), who found that pictures of specific authority figures interacting with subordinates were more effective than the standard TAT in eliciting attitudes toward authority figures.

The specific pictures employed and order of presentation was as follows:

1. Mother and boy with violin (Card 1 of TAT superimposed on Card 5)
2. Employer-employee situation
3. Policeman-subordinate situation
4. Judge-witness situation
5. Buffer: silhouette in window (Card 14 of TAT)
6. Father-son situation (Card 6BM of TAT)
7. Mother-son situation (Card 7BM of TAT)
8. Warden-prisoner situation
9. Teacher-student situation
10. Probation officer-probationer situation

¹ This study was part of a larger interdisciplinary study focused on the interaction process and the sociometric relations developed within a therapy group of probationers to determine whether observational techniques would throw light on reasons for the change in attitude and perceptions of the subjects.

Standard technique was followed (Murray, 1951), for administration before and after treatment. The TAT responses were scored for certain attitudes toward authority figures in accordance with the system devised by Aron (1949) and used by Stern (1952), to wit: (a) Aggression, (b) Independence, and (c) Dependence.

The Human Relations Inventory (HRI), a 37-item projective questionnaire (Bernberg, 1954) which provides a quantitative estimate of social conformity based on the S's need-value system, was also administered before and after treatment.

Independence of treatment and testing was achieved by having an examiner who had no idea of the Ss' degree of verbal participation, administer the tests. Uniform test instructions were given to all Ss and the responses were recorded verbatim. Objectivity in scoring was obtained by using the Aron system which resulted in interjudge agreement of better than 95%. The HRI is completely self-administered and self-scoring.

In addition, the scoring of all tests was completed before the Ss' degree of verbal participation was computed, so that the scorers were not influenced by the latter.

RESULTS

Amount of Verbal Participation and Changes in Attitude toward Authority Figures

The amount of time each S spoke during the fifteen sessions and the scores of the TAT and HRI administered before and after the period of group therapy are summarized in Table 1. The posttreatment score was subtracted from the pretreatment score for each S to obtain a measure of change on the TAT and HRI.

The Ss were ranked according to degree of verbal participation and according to their TAT and HRI change scores. Rank difference correlations were then computed. Rhos of .08 on the TAT and .01 on the HRI were obtained, therefore the null hypothesis that there is no significant correlation between the changes in attitudes and degree of verbal participation cannot be rejected.

DISCUSSION

A common assumption in group therapy is that there is a close correlation between the amount of verbal production and the extent of personal gain derived from the therapeutic experience. It is common practice, for example, in attempting to analyze the "improve-

TABLE 1
NUMBER OF MINUTES OF VERBAL PARTICIPATION AND TAT AND HRI PRE- AND POSTTREATMENT
SCORES OF PROBATION GROUP THERAPY SUBJECTS

Subject	Minutes Verbal Participation	Prescore TAT	Postscore TAT	Improvement TAT	Prescore HRI	Postscore HRI	Improvement HRI
Group I							
A	66	18	16				
B	29	12	11	2	46	40	6
C	63	21	16	1	17	15	2
D	251	9	4	5	20	18	2
E	90	12	6	5	16	16	0
F	88	20	4	6	12	10	2
G	217	9	7	16	16	7	9
H	36	16	11	2	12	6	6
				5	13	12	1
Group II							
I	5	11	8				
J	23	20	10	3	12	8	
K	211	13	8	10	20	10	4
L	195	15	4	5	24	17	10
M	3	14	6	11	12	3	7
N	98	15	5	8	13	7	9
O	107	12	9	10	13	5	6
Means of change				3	19	21	8
				6.13			-2
							4.66

ment" of a member in group therapy, to relate the gross degree of oral output to prognosis. Almost invariably the "silent" member of the therapy group is considered somewhat lacking in reference to improvement. "How could he profit if he said nothing, just listened?" many common sense observers might ask.

Perhaps the important variable which is a function of improvement in treatment is not the torrent of words that emerges from the patient, but the experience of being accepted and understood in the therapy situation, as Rogers (1942, 1951, 1955) has repeatedly suggested. More currently, within the existentialist framework, May (1958, p. 28) has pointed out:

It is well known to every therapist that patients can talk theoretically and academically from now until doomsday about their problems, and not really be affected; indeed, particularly in cases of intellectual and professional patients, this very talking, though it may masquerade under the cloak of unbiased and unprejudiced inquiry into what is going on, is often the defense against seeing the truth and against committing one's self, a defense indeed against one's own vitality. The patient's talking will not help him get to the reality until he can experience something of some issue in which he has an immediate and absolute stake.

On this basis it would appear that the comparatively nonverbalizing patient has an opportunity in group therapy to experience a considerable degree of acceptance and understanding by merely entering the pervasive permissive atmosphere of a functioning therapeutic group without feeling any overwhelming qualms of inadequacy because he is not verbally productive, as might happen in the face-to-face situation of individual treatment.

The anxious therapist, concerned about the group member who does not appear to be actively participating in the verbal interaction, may be reassured that even the relatively silent S may well be employing the therapy period to reintegrate the perceptual

field in the direction of self-realization or he may be staking out an area of commitment. This study at least suggests that the most articulate member of the group is not necessarily the person who gains the most, nor is the quiet one in the group the individual who does not achieve important attitudinal alterations.

SUMMARY

The relationship between degree of verbal participation during group therapy with offenders and their change in attitudes toward authority was investigated. It was found that the amount of verbal production did not correlate with change in attitudes.

REFERENCES

- ARON, B. *A Manual for the analysis of the Thematic Apperception Test*. Berkeley, Cal.: Berg, 1949.
- BERNBERG, R. E. *A Manual for the human relations inventory*. Chicago: Psychometric Affiliates, 1954.
- GIBBY, R. G., STOTSKY, B. A., HILER, E. W., & MILLER, D. R. Validation of Rorschach criteria for predicting duration of therapy. *J. consult. Psychol.*, 1954, 18, 185-191.
- MAY, R., ANGEL, E., & ELLENBERGER, H. F. (Eds.) *Existence: A new dimension in psychiatry and Psychology*. New York: Basic Books, 1958.
- MURRAY, H. A. Use of the Thematic Apperception Test. *Amer. J. Psychiat.*, 1951, 107, 8, 577-581.
- NEWMAN, C. A study of the relationship between attitudes toward authority figures and adjustment to military service. Unpublished doctoral dissertation, New York Univer., 1953.
- POWDERMAKER, F. B., & FRANK, J. D. *Group Psychotherapy*. Cambridge: Harvard Univer. Press, 1953.
- ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.
- ROGERS, C. R. *Counseling and Psychotherapy*. New York: Houghton Mifflin, 1942.
- ROGERS, C. R. Persons or Science? A philosophical question. *Amer. Psychologist*, 1955, 10, 267-278.
- STERN, M. A study of the relationship between attitudes toward certain authority figures and job stability in a group of psychoneurotic veterans. Unpublished doctoral dissertation, New York Univer., 1952.

(Received June 17, 1959)

SYLLOGISTIC REASONING ERRORS IN SCHIZOPHRENIA¹

LEONARD GOTTESMAN AND LOREN J. CHAPMAN²

University of Chicago

Von Domarus (1954) and Arieti (1955a, 1955b) have asserted that an important aspect of schizophrenic thinking disorder is a certain kind of error in syllogistic reasoning. Arieti calls this error tendency "Von Domarus' principle" (or "paleologic") and describes it as follows: "Whereas the normal person accepts identity only upon the basis of identical subjects, the paleologician accepts identity based upon identical predicates" (p. 194). For example, given the premises:

Stags are swift,
Certain Indians are swift

Von Domarus and Arieti both state that on the basis of the common predicate of swiftness, schizophrenics will conclude that "Certain Indians are stags," while normal people will not. Arieti adds (1955b, p. 57) that "Of course this principle of Von Domarus must be viewed as determined by an inability to abstract."

However, other investigators (Chapman & Chapman, 1959; Sells, 1936) have reported strong error preferences of this sort by normals in syllogisms identical in figure and mood³ with the above. Chapman & Chap-

man (1959), using syllogisms containing statements about letters, found that in some moods many college students conclude that two subjects are identical on the basis of a

There are four forms of categorical propositions used in syllogisms:

Expression	Symbol
All pots are black.	A
No pots are black.	E
Some pots are black.	I
Some pots are not black.	O

A syllogism consists of three such statements, the first two of which are the premises.

1. The major premise, which states the relation between the middle term (M) and the predicate (P) of the conclusion.

2. The minor premise, which states the relation between the middle term (M) and the subject (S) of the conclusion.

3. The conclusion, which is the inferred, or deduced, relation between the subject and the predicate.

The figures of the syllogism are the four possible arrangements of the terms in the major and minor premises:

Fig. 1	Fig. 2	Fig. 3	Fig. 4
M-P	P-M	M-P	P-M
S-M	S-M	M-S	M-S
S-P	S-P	S-P	S-P

The mood of a syllogism refers to the combination of three propositions from among the four kinds of categorical propositions. For example, a syllogism of the "AII" mood, in the first figure, is as follows:

All men are mortal.
Some Indians are men.
Therefore:
Some Indians are mortal.

This is a valid syllogism.

Each of the three propositions of a syllogism may be an A, E, I, or O. Considering the major and minor premises only, there are 16 such possibilities of combination. Fourteen of these 16 yield no valid conclusion in one or more of the four figures. The focus of this study is on the latter group.

¹This study was supported by research grant (M-1416) from the National Institute of Mental Health, U. S. Public Health Service.

This article is based in part upon a thesis which was submitted by the senior author to the Graduate School of the University of Chicago in partial fulfillment of the requirements for the degree Master of Arts.

²The senior author is now at the University of Michigan, and the junior author is now at the University of Kentucky.

³The terminology used to describe syllogisms is a uniform one, based on long tradition. For details see any introductory logic textbook, such as Cohen & Nagel (1934). The following summary, however, presents sufficient information for the purposes of this paper.

common predicate, i.e. middle term. For example, in the syllogism

All A's are B's,
Some C's are B's

75% of a college undergraduate sample concluded that "Some C's are A's."

Although both Von Domarus and Arieti cast their examples in the second syllogistic figure, their principle also seems to apply to syllogistic Figures 1, 3, and 4 in the cases in which the premises are logically equivalent to the second figure. For example, II-4 premises are of the form

Some A's are B's,
Some B's are C's.

The premise "Some B's are C's" necessarily means that "Some C's are B's," and making this change in the second premise above would create a syllogism of the second figure. Therefore, the same logical predictions may be made from II-4 as from II-2. In the same way, Von Domarus' principle applies to AA, Fig. 2, AI, Fig. 2 or 4, IA, Fig. 1 and 2, and II, any figure. It appears that their prediction would be an I conclusion for AI and II, and an A conclusion for AA and IA. The latter A prediction is for the premises of the form

Some A's are B's,
All C's are B's.

By Von Domarus' principle, "All C's" and "Some A's" are identical. Therefore, the relationship of C to A is stated by the conclusion "All C's are A's." Previous research (Chapman & Chapman, 1959) found that for premise pairs containing statements about letters, normal Ss chose the Von Domarus predicted error for II, AA, and AI syllogisms. In IA, however, they chose I instead of A.

These findings contradict the claims of Von Domarus and Arieti that such patterns of reasoning are distinctively pathological. Any validity for their reported clinical observations would have to reside in there being an even stronger such tendency in schizophrenics than in normals. This would be consistent with evidence, recently pointed out by one of the present writers (Chapman, 1958) that several kinds of errors which have often been labeled as "the schizophrenic

thinking disorder" are made by both normals and schizophrenics, although less often by normals.

We will compare schizophrenics with normals on the frequency of occurrence of Von Domarus' predicted errors.

If syllogistic errors are important in schizophrenic thinking disorder, it seems unlikely that such errors are limited to these four moods. We will, therefore, compare our groups on all 14 pairings, from among the four forms of categorical propositions, from which no valid conclusion can be drawn. Our main hypothesis is that schizophrenic Ss will exhibit the same error preferences as normal Ss. Since schizophrenic Ss are expected to make more errors than normal Ss, we further hypothesize that their greater number of errors will be expressed as an increased manifestation of normal error patterns.

METHOD

Instrument

The experimental instrument,⁴ a modification of a paper and pencil syllogism test developed by Chapman and Chapman (1959) was given to Ss either individually or in small groups. The test was composed of 40 multiple-choice items, each containing two premises, the word "therefore," and five alternative conclusions. Four of these alternatives were cast in the possible forms which a syllogistic conclusion can take (i.e., A, E, I, and O), and the fifth alternative asserted that "None of these conclusions is proved." For example, the test contained the following item, to which the correct answer is the third alternative which states that no conclusion is proved:

- All of Tom's ties are red.
Some of the things Ada is holding are red.
Therefore:
1. At least some of the things Ada is holding are Tom's ties.
 2. At least some of the things Ada is holding are not Tom's ties.
 3. None of these conclusions is proved.
 4. None of the things Ada is holding are Tom's ties.
 5. All of the things Ada is holding are Tom's ties.

The aim of the present study was to analyze and describe the kinds of errors made by normal and

⁴ Copies of the test may be obtained from the authors, or may be ordered from the American Documentation Institute. Order Document No. 6287, remitting \$1.75 for 35 mm. microfilm, or \$2.50 for 6 × 8 in. photocopies.

schizophrenic Ss on 28 experimental items, which contained two examples of each of the 14 combinations of categorical propositions for which no valid syllogistic conclusion is possible.⁵ The remaining 12 of the 40 items could be solved with one or more valid syllogistic conclusions. These were filler items included to prevent Ss from discovering that the alternative, "None of these conclusions is proved" was the only correct solution on the experimental items. The order of the items and of the alternatives was random.

Preliminary testing of an earlier version of the test revealed that it was so difficult that accuracy scores of normal Ss as well as schizophrenics were close to zero. Such great test difficulty tends to mask differences between groups in accuracy. Therefore, in an attempt to make the test easier, all items were redesigned, so that they had three characteristics. First, in all items the subject, the predicate, the middle terms, and the relationships among them presented objects and relationships that would be within the experience of all Ss, but yet were such that the syllogisms could not be solved on the basis of prior knowledge. That is, the statements all concerned common objects, e.g., toys, automobiles, clothing, but did not specify any particular such objects which the Ss themselves encountered. Second, an attempt was made to construct items in which the correct alternative, "None of these conclusions is proved," was a fairly attractive one. Finally, in the I and O propositions, instead of the word "some" the expression "at least some" was always used, in order to avoid Ss' misinterpreting "some" to mean "some but not all."⁶

In addition to the syllogism test, each S was given the CVS intelligence scale (Hunt, French, Klebanoff, Mensh, & Williams, 1948) as well as the full vocabulary subtest of the Stanford-Binet, Form L.

Subjects

A normal group and a schizophrenic group of 30 Ss⁷ each were tested. Six of the normal Ss were

⁵ Research with this type of instrument by Chapman & Chapman (1959) has shown that the figure of the invalid syllogism has little or no effect on the number or the type of errors made by normal Ss. In addition, none of the writers from whom experimental hypotheses were derived consider the figure of the syllogism. We have, therefore, included examples of only two figures where more than two figures of a given combination of major and minor premise were insoluble.

⁶ See Sells (1936) for a discussion of the ambiguity of the word "some" as it is used in syllogisms. The above change was influenced by his discussion.

⁷ We are indebted to the managers of the West Side VA Hospital, Chicago, Illinois and the VA Hospital, Downey, Illinois for permission to test patients for this experiment. We are also indebted to Meyer Williams and Julian H. Pathman for assistance in obtaining the Ss.

TABLE 1

MEANS OF AGE, EDUCATION, PRESENT IQ SCORE AND HIGHEST WORD PASSED ON VOCABULARY FOR NORMAL AND SCHIZOPHRENIC Ss

	Age	Education	IQ	Highest Word
Schizophrenic	42.3	10.6	89.3	23.9
Normal	38.7	10.9	107.2	25.5

drawn from the general medical and surgical patients at the United States Veterans Administration Hospital, Chicago, Illinois and the remaining 24 from Chicago firemen. No one was included in the normal group who appeared by hospital record or by interview to be mentally defective, to have had a psychiatric or neurological disorder, or to have been a drug addict or an alcoholic. The schizophrenic group consisted of 19 patients at the Chicago State Hospital and 11 at the Veterans Administration Hospital, Downey, Illinois. They had all received a hospital diagnosis of chronic schizophrenia, with their first admission for psychosis occurring at least one year previously and their last at least three months previously. However, almost all had spent most of the last 5 or 10 years in a psychiatric hospital. All but six were closed ward patients, and the majority of them were from wards considered to consist of the hospitals' most regressed patients. Other prerequisites for selection of schizophrenics were absence of any indication of cerebral organic involvement or mental deficiency, no current drug therapy, and no EST treatments within the past three months.

The means of age, education, current functioning IQ, and highest word passed on the Binet vocabulary for the schizophrenic and normal groups are shown in Table 1. The groups were not significantly different on mean of age or education. The schizophrenics were, however, significantly lower on functioning verbal IQ as measured by the CVS. This lower functioning IQ is usually found in schizophrenics because schizophrenia reduces intellectual efficiency. On the variances, the groups did not differ significantly for education and IQ, but did differ for age and more older Ss than the patients. However, we see no reason to suspect that this would affect qualitative error preferences on syllogisms.

Evidence that the groups did not differ in pre-morbid intelligence is seen in the close similarity of the score of highest word passed on the vocabulary. On this score, which is somewhat less subject to lowering by psychosis than is verbal IQ, the two groups did not differ significantly ($p = .78$).

RESULTS AND DISCUSSION

Because the data were extremely skewed and the groups were not large, all evaluations

TABLE 2
PERCENTAGE OF EACH CHOICE ON ITEMS FOR WHICH A AND I ARE PREDICTED
BY VON DOMARUS' PRINCIPLE

Prediction	No. of Items	Group	A	E	I	O	N
A	4	Normal	12	1	34	11	42
		Schizophrenic	20	5	23	20	33
I	4	Normal	6	8	33	23	29
		Schizophrenic	17	10	29	23	22

of differences between normal and schizophrenic groups were made using the Mann-Whitney test. Two-tailed tests were used throughout. Except where specified, comparisons were based on the 28 experimental items.

Both groups erred on a majority of the items, although as expected the schizophrenics made significantly more errors than normals both on the total 40-item test ($p = .03$) and on the 28 experimental items ($p = .05$). The mean numbers of errors of schizophrenics and normals were 20.6 and 17.8 on the experimental items, and were 27.8 and 24.1 for the entire 40 items.

The high number of errors by the normal Ss provides an excellent opportunity for making qualitative comparisons of the preferred errors of the two groups.

Let us first examine the data on those eight items which are relevant to Von Domarus' principle. Table 2 shows the percentage of normal and schizophrenic choices in each error category for the four A prediction items combined (two items with AA premises, both of which were Fig. 2, and two with IA premises of which one was Fig. 1 and one Fig. 2), and for the four I prediction items combined (two items with AI premises of which one was Fig. 2 and one Fig. 4, and two with II, of which one was Fig. 3, and one Fig. 4).

As seen there, schizophrenics qualitatively exceeded normals in Von Domarus' predicted A errors, but the difference was not significant ($p = .31$). In addition, the predicted A choice was not the favored error for the schizophrenics, since O and I alternatives were selected as often as A alternatives. Von Domarus and Arieti seem to predict a con-

clusion of complete identity of subject and predicate (A conclusion) for this kind of syllogism. However, it should be noted that since I is a conclusion of partial identity, the choice of this alternative would not be completely inconsistent with the Von Domarus principle. The A and I choices together account for 46% of the normal Ss' responses on these items, and 43% of the schizophrenics' responses.

On the four items for which Von Domarus' principle would lead to an I conclusion, the schizophrenics did make the largest portion of their errors in the I category but normal Ss exceeded them (not significantly) in this error.

Furthermore, in the case of both sets of items, the schizophrenics and normals showed a quite similar rank order of error preference. It appears that the differences between the two groups on choice of error can be described as a more random response by the schizophrenics. The modal error alternatives are relatively less frequently chosen by the schizophrenics, while the unpopular alternatives are more frequently chosen. A greater scattering of responses by schizophrenics is almost always found on a multiple-choice test.

Another helpful way to view these data is in terms of percentage of errors which coincide with Von Domarus' prediction, i.e., on Table 2 the percentage of the incorrect choices which are A for the four items for which the Von Domarus principle predicts A, and those that are I for the four I prediction items. Grouping together these eight items, the percentage is 35.3% for the normal Ss and 33.7% for the schizophrenics. Contrary

TABLE 3
PERCENTAGE OF NORMAL AND SCHIZOPHRENIC CHOICE OF EACH ALTERNATIVE FOR ITEMS
GROUPED BY NORMAL PREFERRED ERROR

Error Preferred	No. of Items	Group	A	E	I	O	N
O	10	Normal	2	8	18	35	37
		Schizophrenic	11	13	22	29	25
I	8	Normal	8	5	43	16	29
		Schizophrenic	18	7	32	18	25
E	5	Normal	3	30	14	13	40
		Schizophrenic	11	19	24	19	27
A	3	Normal	26	8	16	8	42
		Schizophrenic	22	12	16	23	27

Note.—The categorical proposition form of the premises and the syllogistic figure for which the various errors were preferred were: for O preference, OO-3, OO-4, OI-1, EO-1, IO-3, AI-2, AO-1, IE-3, IO-4, EO-2; for I preference, II-4, II-3, IA-1, IA-2, AO-3, AI-4, OI-3, OA-4; for E preference, AE-1, AE-3, IE-1, OE-4, OE-1; for A preference, AA-2, OA-2, EE-2. These error preferences are more variable than those found in studies (Chapman & Chapman, 1959; Sells, 1936; Woodworth & Sells, 1935) which used relatively content-free propositions, i.e., statements about letters. It is clear that the content as well as the figure and mood influenced the error preferences of the present study.

to the Von Domarus principle, schizophrenic and normal Ss are remarkably similar on this measure.

We conclude that the data do not support Von Domarus' and Arieti's description of differences in syllogistic errors between schizophrenics and normals.

Let us now compare normal and schizophrenic error preferences on the entire 28 experimental items. Items for which the normals had the same error preference were combined for purposes of the comparison. Two items were dropped here because two of their alternatives were tied on frequency of choice. Using this pooling of items, the percentage of normal and schizophrenic choices for each of the five alternatives was computed. These findings are shown in Table 3.

As seen in Table 3, the error patterns for the two groups are similar. Within the pools of items the schizophrenic and normal groups tended to prefer the same erroneous alternatives, and also tended to agree on their least frequent choice, with the differences between the groups of Ss consisting largely of a more uniform dispersion of responses for the schizophrenics. This pattern is marked for

the first two sets of items of Table 3, i.e., those with I and O preference. However, there is poorer agreement on the modal error choice in the case of E items and A items. We find no rationale for the discrepancies which are observed and we do not know if they are reproducible.

An inspection of Tables 2 and 3 indicates that the schizophrenics usually prefer A conclusions more often than the normals, regardless of the premises. This difference approaches significance ($p = .06$) for the pooled 28 experimental items, as computed by a two-tailed Mann-Whitney test. However, this difference may be an artifact of the apparently greater randomness of the schizophrenics' responses, coupled with the fact that the A conclusions are the least popular of the five alternatives for the normal Ss. Support for this interpretation is seen in the fact that on the three items for which the A alternative was the most popular for the normal Ss, the schizophrenics actually made slightly (not significantly) fewer A choices than the normals.

In interpreting the results of an experiment such as the present one it is important to

consider the appropriateness of the instrument. It might be objected that a multiple-choice syllogism test is too artificial a situation to elicit the kind of thinking distortions which are called forth by real events. In particular it might be objected that Von Domarus' principle appears only in affect-laden situations. However, this restriction does not appear to hold for some other aspects of thinking disorder. Several previous studies have amply demonstrated that many clinical thinking disorders can be measured by contrived multiple-choice tests, e.g. Chapman (1956a, 1956b), Epstein (1953), Fey (1951), Gorham (1956), Payne & Hirst (1957). We therefore conclude that the failure of the present study to substantiate the Von Domarus principle indicates that this principle may not be a *pervasive* feature of schizophrenic thinking.

Yet there are occasional schizophrenic deviances which appear to be clear-cut manifestations of Von Domarus' principle, of a sort that normal Ss would not commit. The results of the present study do not contradict such observations, but do indicate that this kind of error may be limited in its occurrence.

SUMMARY

Normal and schizophrenic Ss were given a 40-item multiple-choice syllogism test in order to compare them on the erroneous conclusions which they reach from premises which, by traditional logic, yield no conclusions whatever. These included 8 items which measured the tendency to conclude the identity of subjects on the basis of identical predicates, an error which Von Domarus (1954) and Arieti (1955) have described as distinctively schizophrenic.

The results failed to substantiate the predictions inferred from Von Domarus and Arieti. The present investigators' hypothesis was partially confirmed in that the schizophrenics showed qualitatively similar error patterns to the normal Ss, but was not com-

pletely confirmed in that the schizophrenics did not exceed the normal Ss in these error preferences.

REFERENCES

- ARIETI, S. *Interpretation of schizophrenia*. New York: Robert Brunner, 1955a.
- ARIETI, S. Some aspects of language in schizophrenia. In H. Werner (Ed.) *On expressive language*. Worcester: Clark Univer. Press, 1955b.
- CHAPMAN, L. J. Distractibility in the conceptual performance of schizophrenics. *J. abnorm. soc. Psychol.*, 1956a, 53, 286-291.
- CHAPMAN, L. J. The role of type of distracter in the "concrete" conceptual performance of schizophrenics. *J. Pers.*, 1956b, 25, 130-141.
- CHAPMAN, L. J. Intrusion of associative responses into schizophrenic conceptual performance. *J. abnorm. soc. Psychol.*, 1958, 56, 374-379.
- CHAPMAN, L. J., & CHAPMAN, JEAN P. Atmosphere effect re-examined. *J. exp. Psychol.*, 1959, 58, 220-226.
- COHEN, M. R., & NAGEL, E. *An introduction to logic and scientific method*. New York: Harcourt Brace, 1934.
- EPSTEIN, S. Overinclusive thinking in a schizophrenic and a control group. *J. consult. Psychol.*, 1953, 17, 384-388.
- FAY, ELIZABETH T. The performance of young schizophrenics and young normals on the Wisconsin Card Sorting Test. *J. consult. Psychol.*, 1951, 15, 311-319.
- GORHAM, D. R. Use of the Proverbs Test for differentiating schizophrenics from normals. *J. consult. Psychol.*, 1956, 20, 435-440.
- HUNT, W. A., FRENCH, ELIZABETH G., KLEBANOFF, S. G., MENSCH, I. N., & WILLIAMS, M. The clinical possibilities of an abbreviated individual intelligence test. *J. consult. Psychol.*, 1948, 12, 171-173.
- PAYNE, R. W., & HIRST, H. L. Overinclusive thinking in a depressive and a control group. *J. consult. Psychol.*, 1957, 21, 186-188.
- SELLS, S. B. The atmosphere effect: an experimental study of reasoning. *Arch. Psychol.*, 1936, 29, 3-72.
- VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasanin, (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. Calif., 1954. Pp. 104-113.
- WOODWORTH, R. S., & SELLS, S. B. An atmosphere effect in formal syllogistic reasoning. *J. exp. Psychol.*, 1935, 18, 451-460.

(Received June 24, 1959)

BODY IMAGE CHANGES ASSOCIATED WITH PERSONALITY REORGANIZATION

SIDNEY E. CLEVELAND

Veterans Administration, Houston, Texas and Baylor University College of Medicine

In order to place the present study in proper perspective, it is necessary to summarize briefly the results of two previous research efforts by the author.

First, in a study by Cleveland and Fisher (1958) a relationship between Barrier and Penetration scores and degree of personality reorganization following psychotherapy was found. Patients rated as improved following psychotherapy tended to show increase in Barrier scores and/or decrease in Penetration scores compared to pretherapy scores. Unimproved patients tended to reverse this trend. These preliminary results were tentatively interpreted as suggesting that rated improvement in adjustment paralleled greater definiteness in body image boundaries.

The body image indices of Barrier and Penetration are derived from a content analysis of the Rorschach. In this scoring system a count is made of all responses referring to an object with well defined boundaries, such as "a cave with rock walls," or "a knight in armor." The total number of such responses constitutes the Barrier score. Similarly in obtaining the Penetration score, a count is made of all responses in which the boundaries of an object are weak and ill defined. Such responses as "an open, bleeding wound" or "soft cotton candy" serve as examples.

Results of extensive research indicate that Barrier score reflects degree of definiteness of body image boundaries. Definiteness of boundaries in turn is related to such personality variables as level of maturity and toleration of stress. Penetration score indicates vulnerability of body image boundaries and appears to be negatively correlated with maturity and toleration of stress.

A second report by the author (1958) relates to body image in schizophrenia. In a

normative study, Barrier and Penetration scores were tabulated for 50 normals, 40 neurotics, 40 undifferentiated schizophrenics and 40 paranoid schizophrenics. Normals and neurotics did not differ significantly on either body image scale. However, both groups differed significantly from the two psychotic populations, with schizophrenics having lower Barrier scores and much higher Penetration scores than either normals or neurotics.

This latter study of course represents simply an examination of a cross section of patients hospitalized for a psychotic condition. In order to examine more fully the relationship between body image and ego regression or reorganization, it was desired to follow a group of psychotic patients longitudinally as they underwent treatment. The problem to be studied involved repeated sampling of body image fantasies as the same patient either improved or regressed further over a period of time.

The opportunity to gather such data was provided in the spring, 1958, when the VA conducted a nationwide research project to investigate the relative effectiveness of five different tranquilizers and a placebo (phenobarbital) in bringing about remission of symptoms in schizophrenia. The study was conducted in the usual double blind manner. Participating hospitals included the Houston and Waco VA installations.

PROCEDURE

Twenty-four, newly hospitalized patients with the basic diagnosis of schizophrenic reaction, comprised the Waco sample. Eighteen patients similarly diagnosed were included in the Houston study. Patients to be placed on the research project were carefully selected and assigned by the local project director¹

¹ I am indebted to Donald Gorham for collection of the Waco data.

and psychiatrists concerned. They were guided in their selection by a list of criteria. Among these were the exclusion of patients with lobotomy, CNS disease, and primary alcoholism. Selected patients varied in age from 25 to 45 years and in education from 6th grade to college graduate. All were males and 10 patients were Negroes. Prior to being placed on an unknown medication, behavioral ratings were made of each patient utilizing the Lorr Multidimensional Scale for Rating Psychiatric Patients (1953). This rating was accomplished by a psychiatric team composed of psychiatrist, psychologist, and psychiatric nurse in interview with the patient. Additional ratings were made by the psychiatric aides. These ratings were repeated during the 5th week of drug administration and again following 13 weeks of medication.

The Lorr scale yields an over-all morbidity score. This score represents the degree of pathological behavior observed. Disharmony of thought and feeling, presence of delusional thinking or hallucinatory activity, appropriateness of motor behavior, and degree of reality orientation all contribute to this rating of pathology. The scale is so constructed that these ratings can be made on observable behavior or readily inferable symptoms. Change in morbidity score from first to second and first to third rating serves as one criterion of degree of personality reorganization in the present study. A second criterion of personality reorganization was provided by the psychiatrist's independent global evaluation of the patient made at the conclusion of the study (improved and discharged or unimproved and still hospitalized).

The Holtzman Ink Blot Test (1958) was also administered to each patient at these same three periods during the study: predrug, 5th week, and 13th week. The records were scored for Barrier and Penetration in the usual manner. The Holtzman Ink Blots were utilized rather than the Rorschach test because two carefully equated forms of 45 cards each are available. These were administered in ABA or BAB sequence. Another reason for using this ink blot series was the fact that the subject (S) is required to give only one response per card. Holtzman contends that this system results in a more uniform total response pattern from S to S.

Because of various factors it was not possible to obtain repeat tests and ratings on all Ss. Some patients refused to respond to the ink blot test, either initially or on repeat testing; a few patients were discharged from the hospital before repeat testing could be obtained; one patient committed suicide. However, no one selective factor appeared to be operating to determine which patients would cooperate by responding initially or on repeat testing. For example, some patients rated as improved refused to cooperate on repeat tests, as well as those whose condition had deteriorated.

Overall at least one repeat testing was obtained on 12 of the 18 patients in the Houston group and 13 of the 24 patients comprising the Waco sample. Morbidity ratings based on the Lorr scale were

also available on these 25 patients. Of these 25 patients 17 were given morbidity ratings with significant decrement compared to their initial morbidity score by the 5th week of medication. Eight were unchanged or had increased in morbidity score. By the 13th week 19 patients were given significantly lower morbidity scores as compared with the predrug rating and 6 remained unchanged or worse. Seven of the 25 patients left the hospital on trial visit soon after the 5th week to go back to work. By the end of the experiment 17 of the patients had left the hospital as discharged or on trial visit. These figures are quoted in order to give some perspective on the degree of change apparent in the clinical picture of these patients.

Analysis of the data proceeded in the following manner. The 25 patients available for repeated rating and testing were ranked in order of (a) decrement in morbidity, (b) decrement in Penetration score, (c) increment in Barrier score for the predrug vs. 5th-week and 13th-week evaluations. The patient with the highest decrement in morbidity score was given a rank of 1; the patient with the highest increment in morbidity (much sicker) ranked 25th. For the body image scores, the patient with the greatest drop in Penetration score ranked one and the patient with the largest increase in Penetration score ranked 25th. For the Barrier score ranking the order was simply reversed.

Although Holtzman feels that one advantage of his ink blot series is that comparability of total productivity is encouraged, in the present study patients varied considerably in their response totals from test to test. Accordingly it was found necessary to treat the body image scores as ratios in relation to the number of responses given.

HYPOTHESES AND RESULTS

In view of the previous findings relating change in body image scores to personality reorganization, two predictions were made: (a) a positive correlation was expected between decrement in morbidity and decrement in Penetration score; (b) a positive correlation was also predicted between decrement in morbidity and increment in Barrier score.

A significant rank order correlation of .60 was obtained between morbidity ranking and Penetration ranking for the predrug vs. 5th-week evaluation. Similarly a rho of .61 was obtained on the same rankings for the predrug vs. 13th-week evaluation. These findings support Hypothesis 1.

For the Barrier score and morbidity rankings, however, a rho of .13 for the 5th week and a rho of $-.02$ for the 13th week were obtained. Neither correlation attains statisti-

TABLE 1

A COMPARISON OF PENETRATION AND BARRIER SCORES AND HOSPITAL DISPOSITION

	Penetration Score Decrease	Penetration Score Increase	Barrier Score Decrease	Barrier Score Increase
Patients discharged or on trial visit	15	2	8	9
Patients retained in hospital	3	5	4	4
χ^2 ^a	4.7		0.0	
<i>P</i>	.02, .05			

^a With Yates correction applied.

cal significance and Hypothesis 2 is not supported.

The second criterion of degree of personality reorganization, the psychiatrist's independent global rating of each patient, yielded similar results when compared with changes in body image scores. In Table 1 a comparison by means of chi square technique is made between body image scores and patient disposition by the conclusion of the study. It will be noted that patients considered sufficiently in remission to warrant hospital discharge or trial visit tend to show a decrease in Penetration score. A similar comparison with Barrier score yields a chi square of zero. Again a relationship between Penetration score and degree of personality reorganization is found but no such relationship holds for Barrier score.

Before discussing these findings, two other bits of data should be presented. First it is of interest to note qualitatively the nature of the decrement in Penetration score associated with rated improvement. According to the scoring scheme developed by Cleveland and Fisher for Penetration responses, not only are fantasies involving literal penetration of the body boundary included, but also many responses only symbolic of boundary indefiniteness are so scored. For example: "a shadow"; "soft mud"; "a transparent window." Previous research suggested that acutely disturbed schizophrenics tend to concentrate on literal fantasies of boundary disruption while excluding more symbolic representations. In the present study the patients

gave an average of 9 responses involving literal disruption of the body boundary, human or animal, at the predrug or acutely disturbed period of testing. On later testing those patients rated as improved averaged only 1.4 responses involving literal boundary disruption. For example, one patient initially gave 45 responses involving literal penetration of the body and only 3 such responses on subsequent testing. Another patient dropped from 22 to 5 responses. The following are examples of the kinds of Penetration responses which tended to drop out with rated improvement on subsequent testing: "A man pulled in two by beasts"; "a body torn to pieces"; "bleeding vagina"; "a wound in the skin"; "a scalped head."

It is also of importance to report the reliability of Penetration and Barrier scores in a psychotic population undergoing no significant change in behavior. Opportunity to check the test-retest reliability of these two body image scores was afforded in a study conducted by Rhoda Fisher (1958).² She administered the Rorschach on two occasions five days apart to 50 schizophrenic women. The test was administered during the first week of hospitalization and prior to initiation of any formal treatment program. Test-retest reliability for the Penetration score is .89 and for Barrier score .65. The low reliability correlation for Barrier score is due to the

² I am indebted to Rhoda Fisher for loan of the Rorschach protocols. Scoring of the records and computation of the reliability coefficients were carried out by the present author.

extremely limited range of Barrier responses. For example, Penetration scores range 0-11, while most of the Barrier responses fall within the range 0-4.

It is of further interest that 25 of Fisher's psychotic women were subjected to a stress situation in the nature of a gynecological examination immediately before the initial Rorschach testing. This was done in order to study the differential effect on a number of psychological tests of a situational stress period associated with the testing. In spite of the fact that this examination constituted a literal penetration of the body of these 25 women, Penetration scores obtained from their Rorschach records did not differ significantly from first to second testing nor from those of the control group of 25 women receiving no such stress prior to testing. These findings support the contention that Penetration score represents a basic personality variable involving orientation towards one's body image boundaries and is not merely a reaction to the discomfort of body symptoms or situational stress.

CROSS-VALIDATION STUDY

Opportunity to check on the present findings relating body image scores to rated behavioral change was provided in a study conducted by Goldman (1960).³ Forty-five cooperative, acutely ill schizophrenic patients, ages 16-40, were the Ss of this study. Each patient was administered the Rorschach on admission and again upon leaving the hospital. Each patient was rated twice by two psychiatrists using an 11-item scale. A first rating was done on admission and a second at time of discharge. The psychiatrists agreed that 33 of the 45 cases showed clinical improvement from first to second rating and that 14 of these showed marked improvement. Twelve cases showed no change.

The Rorschach records were scored by the present author for Barrier and Penetration responses. This scoring proceeded on a blind basis as far as knowledge of the psychiatrists' ratings was concerned. Total number of responses was controlled by reducing the re-

sponse total for the larger of the two records to that of the smaller.

When the body image scores were compared with psychiatric ratings, the following differences appeared. For the patients rated as improved ($N = 33$), a mean Penetration score of 3.97 was obtained on first testing, and 1.97 on second testing. A t test for difference in means of related measures yields a value of 3.45, significant beyond the .01 level. Markedly improved patients averaged 4.18 initially on Penetration and 1.91 at discharge ($t = 7.85$, significant beyond .01 level). Unimproved patients also reduced Penetration scores slightly from first to second testing but not to a significant degree. For Barrier, the only significant change occurred in the markedly improved group who raised their mean score from 1.82 to 3.55, first to second testing ($t = 7.52$; $P = 7.01$).

This independent study supports the original findings relating decrement in Penetration score to rated behavioral improvement. An additional finding hypothesized but not substantiated in the first study, concerns significant increment in Barrier score being related to marked clinical improvement. These results are interpreted as lending further support to the idea that definiteness in body image parallels degree of personality organization.

DISCUSSION

Findings in the present study are not reported with any claim concerning degree of recovery or "cure" in schizophrenia. No inference is intended that in 5 or 13 weeks, despite a reduction in morbidity ratings, a cure or complete remission in schizophrenia has been effected. Some of the patients discharged from the hospital at the conclusion of the study have been rehospitalized with a recurrence of symptoms. However, the present concern is not with the therapeutic results of the study nor with the relative effectiveness of different tranquilizers. The focus for the present study is on the nature of the shift in certain aspects of body image within the same individual as he proceeds through his psychosis.

As has been shown, the major change occurring is a dramatic firming up and defining

³ I am indebted to Rosaline Goldman for loan of her Rorschach data and behavioral ratings.

of the body image boundary. This clarification of the boundaries takes the form of a sharp reduction in fantasies involving disruption and dissolution of the body image boundaries. Acutely ill schizophrenics present a concentration of morbid, wildly disrupted fantasies portraying a literal breakdown of the body boundaries. Under a therapeutic regime these bizarre fantasies very quickly disappear, indeed, at least as early as five weeks after introduction of medication.

Alteration in the boundary aspect of body image in schizophrenia is a finding in keeping with a number of studies. Szasz (1957) observes that the profusion of bodily feelings in schizophrenia expresses "the inner experience of personal dissolution at the periphery. . . ." Szasz has enumerated a series of phases in the genesis of bodily feelings in schizophrenia. Of interest to the present paper is his observation that in the acute and active period of a psychotic break, bodily preoccupation is focused largely on the loss of the body boundary as an object to which to relate. In terms of the present study it is conceivable that administration of tranquilizers and closed ward hospitalization serve to provide for the patient a sharper delimitation of body image from environment.

Other investigators emphasizing the role of body image boundaries in schizophrenia include Bender & Keeler (1952) who have studied body image formation in schizophrenic children via the medium of figure drawings. They conclude that schizophrenic children "show in particular a difficulty in determining the periphery of their own body or the boundaries of their personality or 'ego boundaries.' There is an indefiniteness and blurring of the zone between their own body and the outside world." These conclusions were drawn largely from the gross distortions of various body parts introduced into the drawings by the children. Interestingly, in comparison with the present findings concerning Penetration responses, Bender and Keeler noted especially that "body orifices were accentuated and displaced and body boundaries received undue treatment" in the children's drawings.

Lastly, Weckowicz and Sommer⁴ have explored body image in a schizophrenic and nonschizophrenic population. They find that schizophrenics consistently estimate pictures of hands and feet to be significantly smaller than do nonschizophrenics. Schizophrenics also make significantly smaller self drawings of hands and feet when viewing themselves in a mirror than do nonschizophrenics. The groups did not differ in size estimation or drawings of central portions of the body. The authors attribute this perception by schizophrenics of distal parts of the body as smaller to a shrinking of the periphery of the body image.

The importance of establishing a well defined body image boundary as a prelude to more total ego reorganization has long been recognized by some treatment programs. For example, Reich (1949), Bettelheim (1950) and others have formalized the importance of the reinforcing of body limits in the treatment of schizophrenic and severely neurotic patients. Programs of physiotherapy, hydrotherapy, and cosmetics all implicitly reinforce the patient's body periphery by providing maximal stimulation to this area. Again what is emphasized is the value of demonstrating to the schizophrenic the reality and existence of his own body boundaries. With the return of the substance of his body periphery, the schizophrenic can then generalize this body reality to the area of ego reality.

SUMMARY

Twenty-five recently hospitalized schizophrenics were administered the Holtzman Ink Blot Test and Lorr scale prior to being placed on an unknown medication. These measures were repeated after 5 and 13 weeks on medications. The ink blot test was scored for the body image indices of Barrier and Penetration. A morbidity score obtained from the Lorr scale ratings constituted one criterion of degree of personality reorganization. Psychiatrist's disposition of the patients

⁴ Personal communication. Weckowicz, T. E. and Sommer, R. Body image and self-concept in schizophrenia. Unpublished paper, 1959.

by the end of the study (discharged vs. still hospitalized) constituted a second criterion.

A rho of .60 between decrement in Penetration score and decrement in morbidity for the 5th-week evaluation was obtained. A rho of .61 was maintained for the same rankings at 13 weeks. Increment in Barrier score was not found related to morbidity ranking. Similar results were obtained for the second criterion measure, psychiatrist's disposition of the patient vs. body image scores. Implications of these results for the nature of body image changes in schizophrenia are discussed.

REFERENCES

- BENDER, L., & KEELER, W. R. The body image of schizophrenic children following electro-shock treatment. *Amer. J. Orthopsychiat.*, 1952, 22, 335-355.
- BETTELHEIM, B. *Love is not enough*, Glencoe, Ill.: Free Press, 1950.
- FISHER, R. L. The effect of a disturbing situation upon the stability of various projective tests. *Psychol. Monogr.*, 72(14, Whole No. 467).
- FISHER, S. & CLEVELAND, S. E. *Body image and personality*. Princeton: Van Nostrand, 1958.
- GOLDMAN, R. Changes in Rorschach performance and clinical improvement in schizophrenia, *J. Consult. Psychol.*, 1960, in press.
- HOLTZMAN, W. H. *The inkblot test*. Austin, Tex. Univer. Texas Press, 1958.
- LORR, M. Multidimensional scale for rating psychiatric patients, hospital form. *VA tech. Bull.*, 1953, TB 10-507.
- REICH, W. *Character analysis*. New York: Orgone Inst. Press, 1949.
- SZASZ, T. S. *Pain and pleasure*. New York: Basic Books, 1957.

(Received June 29, 1959)

FACTOR ANALYSES OF THE RORSCHACH¹

BERNARD I. MURSTEIN

University of Portland

Despite the plethora of studies utilizing the Rorschach (currently well past the 2,000 mark) the amount of knowledge obtained has been disappointing. The reasons for this unfortunate circumstance have been the refractoriness of the technique to *crucial* experimentation. While tens of iconoclastic researchers have hammered away at the folklore of the Rorschach (e.g., the number of responses to Cards VIII, IX, X reflect emotional reactivity) the "defenders" have retorted that current statistical methods are insensitive to the important configurational and test-interpretive aspects of the instrument.

The increasing importance of factor analysis in psychology made it inevitable that it would be utilized in research involving the Rorschach. It is the purpose of this paper to assess the contribution that factor analysis has made with regard to this technique.

The problems inherent in the use of factor analysis will be dealt with first. Next the difficulties peculiar to the use of the Rorschach as a psychometric instrument will be discussed, followed by a section on the various Rorschach hypotheses tested by means of factor-analytic methodology. Then, a section listing the major new contributions to the Rorschach will be presented, followed finally by a discussion of the newer Rorschach innovations which have attempted to create an instrument more amenable to psychometric treatment. A summary of each factor analysis involving the Rorschach only is presented in Table 1. Studies in which the Rorschach was investigated in addition to other variables are summarized in Table 2. These tables indicate the population sampled,

the number and kinds of variables studied, the types of correlations used, and the factors obtained.

PROBLEMS INHERENT IN FACTOR ANALYSIS

In a study of Yale undergraduates, Wittenborn (1950a) used the factor-analytic method to test hypotheses. Thus, he stated four hypotheses and attempted to substantiate them via his rotated factors. Whether this is a valid test is debatable since solutions with or without rotation are not mathematically unique. There are many possible rotations and one may rotate the axes infinitely until one gets them into the position which best substantiates the hypothesis.

It would seem that an "objective" means for evaluating his hypotheses would have been through "blind rotations" (i.e. rotation according to some criterion such as "simple structure," but without knowledge of the nature of the variables involved). It is not apparent from the report that Wittenborn followed this procedure.

Among the assumptions made in the use of factor analysis, two are especially cogent with reference to the Rorschach. These are: (a) the variables must be linearly additive, and (b) the variables must be formally independent. A violation of the latter assumption occurs in the work of Adcock (1951). He treats as independent scores the ratios $W : M$ and $M : C$. The common element in both (M), assures that if the former ratio is low due to the presence of many M s the latter ratio must be high because M is now in the numerator of the second ratio as compared to its being in the denominator of the first ratio. Another example of linear dependency is the use of percentage scores. Hence, a high negative loading for $W\%$ virtually assures a negligible or high positive loading for $D\%$ or $Dr\%$.

¹ The author would like to express his appreciation to William A. Botzum and Walter G. Klopfer for generously giving of their time in reading the manuscript and offering their criticisms.

TABLE 1
SUMMARY OF FACTOR-ANALYTIC STUDIES WITH THE RORSCHACH

Investigator	Subjects	No. of Variables Analyzed	Type of Correlation	Factors Extracted and Solution
^a Hsu (1947)	76 emotionally disturbed children, ages 5-15	15	Tetrachoric	Orthogonal Facility in use of words Facility in use of nouns "Face" factor Facility for verbs and adjectives "Human" factor
^b Wittenborn (1949)	247 Yale undergraduates	18	Tetrachoric	Orthogonal Six factors Oblique Seven factors; interpretation of factors is largely avoided
Wittenborn (1950a)	92 Yale undergraduates	21	Pearson	Orthogonal Productivity Low perceptual control Empathy One unnamed factor
Wittenborn (1950b)	160 in- and out-patients from psychiatric clinic	21	Pearson	Orthogonal Bizarre originality Low perceptual control Productivity Empathy ?
Adcock (1951)	88 Cook Island children 30 New Zealand children	15	Tetrachoric	Cook Island Ss: Oblique Fluency Introversion Intelligence Construction New Zealanders: Orthogonal Fluency Introversion Intelligence
Coan (1956)	Wittenborn (1950a) 92 Yale undergraduate data	12	Pearson	Orthogonal Productivity Low perceptual control (anxiety) Low perceptual control (Passive submission to emotional impulses) Intratensivity Inner control and empathy Outer control
Stotsky (1957) Location	148 Schizophrenics	3	Tetrachoric	Orthogonal (location scores) Whole Common and rare details
Determinants	148 Schizophrenics	5		Orthogonal (determinant scores) Form Color and shading

^a Only Card I administered.

^b Group administration (Harrover-Erickson check list).

TABLE 2
 RORSCHACH INCLUDED AMONG OTHER TESTS

Investigator	Subjects	No. of Variables Analyzed	Type of Correlation	Variables	Factors Extracted and Solution
Sen (1950)	100 Indian College students in England	38	Tetrachoric	Verbal Intelligence, Non-verbal Intelligence and 36 Rorschach variables	Orthogonal Fluency Cognitive (synthetic vs. analytic) Emotional (neurotic vs. non-neurotic) Group factors Fluency (basic factor) General intelligence General emotionality Reproductive association
Hughes (1950)	100 psychiatric hospitalized patients	22	Tetrachoric	22 signs of organicity from Rorschach	Orthogonal Organic factor; factors 2 through 8 not labeled.
Cox (1951)	120 (60 normal school boys, 60 child guidance clinic boys)	27	Tetrachoric	School vs. clinic category, IQ, 25 Rorschach variables	Orthogonal Productivity Normal vs. maladjustment Intelligence vs. lack of intelligence Initiative—Passivity Accurate-inaccurate perception
Sandler and Ackner (1951)	50 psychiatric patients	50	?	50 psychiatric patients (Modified Q analysis)	Orthogonal High vs. low productivity Perception of internal, anatomical objects vs. perception of external objects Animate percepts vs. inanimate percepts Defensive percepts vs. well-defined human parts
Lotsof (1953)	30 college students	11	Pearson	Intelligence test, total words, adjectives, verbs, 7 Rorschach determinants	Orthogonal Verbal intelligence Productivity Elaboration Individuality
Williams and Lawrence (1953)	100 psychiatric patients	22	Tetrachoric	Verbal IQ, Performance IQ, 20 Rorschach determinants	Orthogonal Productivity Lack of perceptual control Intelligence Movement Shading
Williams and Lawrence (1954)	100 psychiatric patients	32	Pearson and Tetrachoric	21 Rorschach 11 MMPI	Orthogonal and Oblique Rorschach productivity Ego strength Expressive-Repressive MMPI maladjustment
Foster (1955)	1. 54 Policemen	11	Tetrachoric	Vocabulary, Verbal IQ, Anxiety ratings, 8 Rorschach variables	Factors identified from two matrices on the basis of similar loadings on common items (proportional profiles)
	2. 28 College students	19	Tetrachoric	Allport-Vernon Scale of values 6, MMPI 6, Wechsler 3, Rorschach variables 4.	Orthogonal Inhibition of spontaneity Verbal intelligence Emotional drive Anxiety Manic-depressive
Borgatta and Eschenbach (1955)	125 Air Force personnel	40	Pearson	Rorschach variables 15, Interpersonal performance scores 11, rating scales 7, PMA 3, Actuarial 4.	Orthogonal Rorschach productivity Formal properties Rorschach Abstract intelligence Social intelligence Emotional assertiveness Task leadership Social acceptability Maturity-adjustment

TABLE 2 (Continued)

Investigator	Subjects	No. of Variables Analyzed	Type of Correlation	Variables	Factors Extracted and solution
Singer, Wilensky & McCraven (1957)	100 male veteran schizophrenics	23	Phi coefficient	Rorschach variables 7, Wechsler-Bellevue IQ, Barron's Movement Threshold Inkblots, TAT Transcendence Index, Planning Self-Ratings, Porteus Mazes, Wechsler's Number Square test 3, Time Estimation, Block Authority Reaction, Motor Inabilities Test, Digit Frustration, Behavior Ratings 4.	Oblique Motor inhibition and planfulness Ambitiousness Emotional surgency Introspectiveness
Consalvi & Center (1957)	45 adults (6th grade education to professional college)	14	Pearson	Raven progressive matrices, Vocabulary, 12 Rorschach Variables	Orthogonal Intelligence Low form Productivity Movement
Lotsof, Comrey, Borgartz, & Arsfield (1958)	72 underachieving school children	33	Pearson	Actural Variables 3, Verbal IQ, Performance IQ, Wechsler subtests 12, Rorschach variables 16.	Orthogonal Verbal Intelligence Productivity Perceptual movement nonhuman Age Performance IQ Seven unnamed factors
Vernier, Stafford, & Krugman (1958)	150 physically disabled male war veterans	35	Tetrachoric	Otis S. A. test of mental ability, Bender-Gestalt 4, Rorschach 12, Draw-A-Person 6, Sentence Completion 7, Physical pathology 5.	Oblique Withdrawal Tension Need for social approval Anxiety Dependency-Passivity Affective Liability Pseudo-Intellectual Defensiveness

The variation in methods of rotation also has doubtlessly contributed to the emergence of different factors. While the centroid solution with either orthogonal or oblique rotation has been favored, other variations have been used. The British analyses have included Burt's simple summation and group methods. Other variations have included the Kaiser varimax method and various principle components solutions. The use of different methods contributes somewhat to the differences in solution.

Another problem is the fact that there is currently no method of obtaining a standard error of a factor loading. Hence the stability of the position of each factorial loading can be only indirectly assessed via such criteria as the number of subjects (Ss) used and the type of correlation employed. There are ways of reducing the risk of error by the employment of techniques such as repetition of previously found factors, rotation to known "psychological conditions" and the use of

"marker" variables whose factorial makeup is known. These aids, however, are often not present in the first factorial investigation of a test.

PROBLEMS SPECIFIC TO THE RORSCHACH

Nonstatistical Problems

While the Rorschach is ostensibly a *perceptual test*, operationally speaking it must be considered a *verbal task* only indirectly getting at the perception via the verbalization. This fact may not cause any consternation until one reflects fully upon the fact that the testee must manifest an extreme sensitivity to the determinants of his perception. He must communicate the importance of form, and in the case of using shading, whether the concept is two- or three-dimensional, whether it involves the perception of "texture" or perhaps "vista" to mention just a few possibilities. But even if a person possesses a keen awareness of the determinants of his perception, does he pos-

sess the vocabulary with which to communicate them to the examiner? It is apparent that to convey verbally the nuances of his perception a person needs a precise and extensive vocabulary. One may well wonder whether the dearth of determinants found in low socio-economic class protocols reflects a "simple" perceptual world or a small and limited vocabulary.

The fact that a few "incidental" words may change the scoring of a determinant is yet another difficulty. If two men are seen on Card III the determinant might be an *F*. But, if on inquiry the two men are "sitting down," the response is scored not *F*, but *M*. The examiner's task is to determine whether the original percept involved movement but was not verbalized, or whether the movement really was not a kinesthetic perception but a verbal embellishment.

The fact that a single perception may be subject to many different scorings, depending upon the verbalization of the *S*, is another weakness in the test as Levin (1953), Baughman (1954, 1958, 1959), and Murstein (1958) have pointed out. This point is amply demonstrated in a study by Stotsky (1957) where the determinants *D* and *F* loaded .93 and .96 respectively on the same factor. If almost every percept scored *D* for location is also form-determined (*F*), then they can hardly be conceived of as independent variables.

The same situation holds for the treatment of certain determinants and content scores as independent items. Can there be human movement (*M*) by other than a human being (*H*)? In a small number of cases the answer is yes, e.g., animals perceived as "dancing a polka." In most instances, however, the number of *M* correlates .90 or above with the number of *H* + *Hd*. Such a correlation then involves little psychologically meaningful covariation but is in essence an artifact. Its presence in the factor matrix may lead to erroneous interpretation.

Another problem is that if a perception involves several determinants a specific order of scoring may be used which allows only the determinant highest in the hierarchy to appear to be present more often than the lesser favored determinant. Accordingly, *M*

takes predominance over *C*, which takes predominance over *Fc*, etc. For maximal interpretative value, however, the number of responses should be independent of the frequency of occurrence of the determinants.

Clinicians through the years have grumbled over the necessity of learning at least two systems of scoring (the Beck, and Klopfer system), not to mention the somewhat less popular Hertz and Buhler variations. The minor difficulty of understanding the various Rorschach dialects is simple in comparison to the man-sized task facing the reader of the Rorschach factor analyses. It is common knowledge that factors are not invariant when the population or tests used in the factorial matrix are varied extensively. A brief listing of procedural variations should suffice to indicate that no two Rorschach matrices reviewed herein have been identical:

1. Different tests have been used. The Harrower-Erikson check list, for example (Wittenborn 1950a), is not comparable to the usual Rorschach.

2. Different scoring systems have been used such as the Beck and Klopfer ones, in addition to mixtures in varying strengths of the two.

3. Some analyses include the form level of a given determinant, others do not.

4. Sometimes the whole test has been analyzed, at other times only a part of the test.

5. Sometimes the instructions have been the "usual"² ones while at other times the *Ss* have been told to produce a fixed number of responses.

6. The researchers have differed in their choice of variables to analyze. No two matrices analyzed by different investigators have contained the identical variables.

7. The telescoping of scores has been freely undertaken. Thus, one investigator used three separate color variables, *FC*, *CF*, and *C*; another used only ΣC . Some have analyzed shading and texture as separate scores while

² Their "usual" instructions are "usual" only in a gross sense since as has been brought to this author's attention in a personal communication by Walter G. Klopfer, the Beck system by its instructions implies a much stronger set to produce responses than does the Klopfer system.

another has combined shading and texture into one score.

8. Some researchers have correlated the raw scores; others normalized scores; still others used percentages and ratios.

Statistical Problems

Choice of correlation. The Pearson product-moment correlation is the most frequent statistic used when a measure of linear regression is involved, and several studies have utilized it. It is, however, justifiably applied only when (a) the regression is linear and (b) the variables possess the property of homoscedasticity. It is apparent from common clinical experience as well as such articles as the one by Fiske and Baughman (1953), that the relationship between most of the Rorschach variables is not linear due to the stimulus limitations of the cards which favor the differential occurrence of the determinants.

Some investigators have realized the inapplicability of the Pearson r and substituted instead the tetrachoric correlation. This statistic, however, also assumes normality of distribution and linear regression. In addition, the tetrachoric r has a greater standard error than the Pearson r and tends to overestimate the true correlation. To avoid the dubiousness of using the Pearson r as well as the tetrachoric r , the phi coefficient has been used because it does not assume underlying linearity nor a normal distribution. Its limitations, however, are formidable. It is not an estimate of a parameter, does not always vary from +1 to -1, and has no confidence limits. Another approach is to normalize the raw scores and then correlate the normalized scores using the Pearson r . This approach, however, seems to violate the logic involved in normalizing data. The assumption is usually made in normalizing data that the trait or ability would be distributed normally in the population if we had sensitive enough measurements. It is the nature of the Rorschach itself and not only the crudeness of the quantitative scores which do not justify a normalizing procedure. With the small number of 10 cards and the various stimulus limitations inherent in these cards it is unlikely that the perception of the various determinants can in any way be assumed to be

basically normally distributed. In any event, it hardly could be claimed that the small select populations used in the factor-analytic studies were themselves normally distributed, thus justifying the use of such scores with them.

Yet another problem involved in the use of normalized scores is that they have the further disadvantage of transformations in general, in that they render the original data meaningless, or as Wittenborn has succinctly put it, it "analyses the Rorschach as it isn't." The resulting conclusion is that the use of correlational indices are, strictly speaking, unjustified. The crucial question, then, is how much does one distort the findings in using a given statistic? This question is extremely difficult if at all possible to answer. Until some statistician solves the dilemma one may only conclude that most of the correlational matrices used have had an indeterminate amount of error variance and gave at best only approximate factorial solutions.

Another problem is the nonmutually exclusive character of Rorschach scores. These scores by their nature determine that the correlations between the components must on the average be negative. This is easily seen with the location determinants, W , D , and Dd . If in a given protocol, the total number of responses is 30, and 8 responses are scored W , the limitation is that the remaining responses must total 22. In a series of protocols then there must be a perfect negative relationship between W and $D + Dd$, if R is a constant. That R is not constant from person to person in most studies simply means that this relationship will be attenuated somewhat but still present.

The use of percentages instead of raw scores may often lead to unwarranted conclusions as illustrated in a study by Sen (1950). By means of judges' ratings and Burt's Rorschach content method, a significant correlation was reported between the score "imagination" and the "productivity" factor. Sen noted that M often assumed to tap "inner creativity" had no loading on this factor. This finding is attenuated because $M\%$ was used instead of number of M . Consequently it may well have been that imaginative individuals manifested a high R along

with a high number of M . If, however, the slope of $\Delta M/\Delta R$ is one of negative acceleration as R increases as seems indicated by the data of Fiske and Baughman (1953), the $M\%$ will not remain constant with the increase in number of responses, but rather decrease somewhat. Thus, the failure of $M\%$ to load on the factor of productivity may reflect not lack of imagination but the stimulus limitation for the perception of M on the cards.

Unfortunately, the Rorschach Psychodiagnostic was not conceived by a man familiar with the problems of quantitative measurement. Indeed, Rorschach scores like, M , F , CF , etc., are not scores at all but a shorthand for clinical impressions. By way of comparison, a Wechsler Vocabulary Scale score of 18 indicates that the S is extremely conversant with words, and probably of extremely high intelligence level (r of .85 with Full Scale IQ). That a S has seven M s in his protocol, tells us very little. One wishes to know what the form level is, whether the M s are manifested spontaneously or elicited via inquiry, whether the M s were given on Card III or IV, and whether they refer to a large or small section of the blot. All of these questions are important to the clinician but usually none are reflected in factor analyses of the Rorschach. It is little wonder then that successive analyses with these "scores" seldom are comparable.

Another problem often ignored is the extremely low reliability of many Rorschach scores, including the various color and shading measures. One cannot expect very meaningful factors unless the components which compose these factors, i.e. the correlations, possess adequate reliability.

The problem of R , the number of responses as a variable contributing spurious variance to the factorial structure has been excellently dealt with in a study by Coan (1956). He states that:

The correlation of R with any location or determinant score is actually a part-whole correlation, since the sum of either the primary determinant scores or of the location scores will equal R . Similarly, P and O are functions of protocol length, since their size is limited by the total number of responses and longer protocols present more opportunity for

popular and original responses to occur. On a similar basis, certain specific location and determinant scores may be expected to correlate especially highly with R , so that even percentage scores will show certain systematic correlation effects (p. 282).

As an antidote to this problem, Coan proceeded to analyze his data omitting R . The omission of R , however, does not necessarily remove the spurious component of correlation from any two determinant scores whose correlation is due to their mutual dependence on R . In short, Coan's analysis is only partially free of spuriousness, and accordingly does not reflect the "intrinsic" relationship between the determinants. On the other hand since it has some of the spurious variance removed it does not reflect the "external" relationship either.

Along the same line, Glickstein (1959) recomputed two of Wittenborn's matrices (Wittenborn 1950a, 1950b) with R partialled out. The results showed that the mean of the matrices with R partialled out approximated zero while Wittenborn's original matrices appear to have a mean of about .30. The reply of Wittenborn (1959) indicates that their differences may be due to different assumptions. Wittenborn holds that R is the resultant of the number of responses in the various scoring categories and not the generator of these responses, as Glickstein seems to imply. There is no direct evidence on this point. Webb and Hilden (1955), however, have shown that productivity on the TAT is related to verbal fluency. If it could be demonstrated that Rorschach productivity is related to productivity in other areas (i.e. verbal fluency) the argument that R generates the determinants rather than vice versa, would be greatly strengthened. The origin of R is important since if R is due to a "set" the benefit of partialling it out may be the resulting gain in knowledge about the interrelationship of the determinants. If R is not due to a "set," Wittenborn is correct in stating that we do violence to the data in partialling it out. Until further research is forthcoming, the issue remains in doubt.

Wittenborn (1959) is on firmer ground when he points out that the use of partialling techniques ignores the actual responses given to the Rorschach which are the most mean-

ingful from an interpretative point of view, "At present one has the unhappy choice of studying the Rorschach 'as it is,' of studying it 'as it isn't,' or of ignoring it altogether" (p. 77).

EVALUATION OF RORSCHACH HYPOTHESES

Many factor-analysts have been interested in submitting much of the Rorschach folklore handed down from clinician to clinician, to a factorial test. In the following pages some of their findings will be briefly discussed.

$M : \Sigma C$ (Erlebnistyp)

Wittenborn (1950a) in a study of 92 Yale undergraduates found M to be factorially distinct from C and CF , though not from FC which partially substantiated his hypothesis that M and the color responses would be factorially distinct. Such evidence seems to be detrimental to the continued use of the $M : \Sigma C$ ratio. Ainsworth (1954) in a rebuttal makes much of the fact that the Rorschach categories are not based on a factorial rationale. Most Rorschachers do not treat color unidimensionally. FC is held to be qualitatively as well as quantitatively distinct from CF and C . Further, the relationship of M and FC is not unexpected in the normal adjusted personality. Even from a formal point of view it should be readily apparent that M and FC might have a good deal more in common than FC and CF . The scoring of M invariably implies the presence of form, for it is impossible to conceive of formless humans. It is possible that the movement response also might have included the perception of color although this might not be scored since M takes precedence over C in some scoring schemes. One may therefore conclude that the appearance of M and FC on the same factor probably indicates their mutual emphasis of form as well as their socially adjustive implications. While FC and CF have color in common they differ in the importance of form for the perception, and apparently this latter difference is the more important one since it results in their loading on separate factors. These statements put the Rorschacher in the position of having his cake and eating it too. It should not be possible to defend the act

of distinguishing between FC and CF in one breath and in the next breath proceed to treat both as similar in compiling a $M : \Sigma C$ ratio.

Adcock (1951) investigated the hypothesis that *Erlebnistyp* ("Experience Balance") is indicated by several indices: (a) $M : \Sigma C$, (b) $FM + M : Fc + c + C'$, and (c) number of responses to cards VIII, IX, X. The validation of the hypothesis was held to be dependent on their high loading on a common factor. This did not occur and Adcock concluded that all three indices did not measure *Erlebnistyp*. To some Rorschachers, this conclusion would not be valid. They could argue that $FM + M : Fc + c + C'$ represented a potential ratio of intraverative to extraverative tendencies which were not currently realized. There is no reason, however, why one's potential extraverative tendencies should be a simple function of one's "realized" degree of extraversiveness. The exact opposite might be argued in that if one is realizing much of his reservoir of extraversivity there should be little left in reserve. While this argument weakens Adcock's finding, it does not inspire confidence in the testability of the Rorschach hypotheses.

Intelligence

Many Rorschachers have held that intelligence can be estimated from several Rorschach determinants as well as nonformal scores such as variety of content and originality of percept. Among the formal scores, intelligence is asserted to be present when W , M , R , and F are present, to name just a few scores. The factorial studies have in general supported these hypotheses. M has been highly loaded on the intelligence factors found by Sen (1950), Lotsof (1953), Williams and Lawrence (1953), and Consalvi and Center (1957). W has been highly loaded in three of these studies, with F and R each represented in one study. Only Lotsof, Comrey, Bogartz, and Arsfield (1958) failed to find any of these determinants on their "intelligence" factor. Their study, however, utilized children as S s and the "intelligence" hypothesis has not been generally extended to apply to children.

Neuroticism

The number of "neurotic signs" referred to in the Rorschach literature is legion. It is difficult to assess whether factor analysis aids in validating these signs when we cannot state an operational Rorschach definition of neuroticism. Despite the variation in choice of "neurotic" determinants, however, one category seems to be present on almost all lists. This is the color category, with particular reference to *CF* and *C*. These determinant scores have been used in studies where Rorschach "neurotic" factors were obtained. Sen's "neurotic" factor does contain high loadings for *C* and *C + F* which is somewhat expected in view of the common component *C*. These signs are given substantial support by virtue of their correlating highly with several other "internal" and "external" criteria. The factor was found to correlate approximately .7 with Burt's Rorschach Content Method Neurotic Rating and with peer-judges' ratings for neurotic tendencies. Both *CF* and *C* also had a significant negative loading on an "adjustment" factor reported by Cox.

Lotsof (1953) reported putatively contrary findings. Since ΣC and *Y* (shading) had loadings of .70 and .40 respectively on the same factor, Lotsof concluded that his study had serious negative implications for Rorschach theory since "Anxiety [*Y*] and affect [*C*] should not appear correlated for a 'normal' population to the extent that they do in this study, . . ." (p. 24). By way of retort one may state that without knowledge of what is meant by "anxiety" and "affect" it is hardly possible to say whether or not they are correlated in a "normal" population. Besides, the equating of ΣC to "affect" and *Y* to "anxiety" without any further investigation into the form level and configural aspects of both of these categories is without justification.

The picture is clouded, however, by the results of Williams and Lawrence (1954). Their "ego strength" factor contains moderately high *positive* loadings for *C* and *CF*, whereas the expectation should be for negative loadings.

In sum, the overall picture is unclear as to the determination of any kind of neurotic

sign or validation of *Erlebnistyp*. The evidence seems favorable for the use of Rorschach determinants for intelligence. One vital flaw which would be emphasized by many clinicians is that most of the studies have no reference to the form level of the determinants. This appears to be a valid criticism and it is strange that no experimenter has taken cognizance of this important point. Accordingly, without reference to form level, the use of the determinants cannot constitute an adequate test of Rorschach hypotheses which do emphasize the value of form level in reference to *Erlebnistyp*, intelligence, and neuroticism.

OTHER RORSCHACH FINDINGS

The use of factor analysis has importance for increasing knowledge concerning the Rorschach apart from testing the commonly used hypotheses. The following section lists some of the newer findings:

Productivity Factor

Most of the Rorschach studies have reported a "productivity" factor which reflects the degree to which the manifestation of the various determinants are a function of the total number of responses in the protocols. The importance of this factor for clinician is that it enables them to test hitherto untested hypotheses as well as suggesting new ones. For example, the presence of a large number of *D* has usually been interpreted as indicating a preference for the larger details of a task rather than the organizational aspect. Factor analysis, however, has indicated that *D*, *d*, *Dd*, *F*, and *R* are highly loaded on the "productivity factor." Since the major portion of common variance of each of these determinants is accounted for by this factor it is possible to examine the usual psychological meaning applied to each of these determinants and see whether the simultaneous existence of all of these psychological hypotheses is clinically meaningful. In the aforementioned factor this would mean that the tendency to attend to the larger details of a situation (*D*) is also accompanied by the tendency to attend to small but articulate details (*d*) and to the tiny aspects (*Dd*). In short, we could economize by say-

ing that the tendency to attend to details is general no matter what the size of the detail. But, the productivity factor also indicates that contour determined perceptions (F) and general productivity (R) should be related to the tendency to perceive details. It is apparent then that insofar as meaning exists in determinants codified so as to be amenable to factor analysis, the factors extracted are helpful in testing the validity of psychological hypotheses as to the meaning of these determinants.

Psychiatric Diagnosis

Several researchers have addressed themselves to the question of whether the factorial structure of psychiatric patients could be differentiated from that of normals. The evidence would appear to be negative. Wittenborn undertook factor-analyses of both kinds of populations. The Ss were unmatched, the normal population consisting of 92 Yale undergraduates (1950a) while the psychiatric one was composed of 160 in- and out-patients from a psychiatric clinic (1950b). Four factors were obtained for each group, of which three appear highly similar (Productivity, Low Perceptual Form, and Empathy). The chief difference between the two populations was a splitting of a factor found in the normal population which had high loadings on M , m , Fc , C' , FC and original responses O (presumably reflecting a healthy controlled degree of empathy and originality) into two separate factors, one reflecting the movement and form controlled determinants, the other possessing high loadings on O and S which now reflected a bizarre kind of originality.

The similarity of the factorial makeup for the two populations is surprising. It may well be that the formal determinants investigated in an atomistic way without regard to configural pattern or form level may not be differentiated by mental disease. This might be due to the stimulus properties of the cards, the Ss' test-taking attitudes, or a myriad of other reasons.

Moreover, the meaningfulness of the factors is difficult to assess because of the lack of external criteria. Ainsworth (1954) correctly points out, for, example, that rather than reflecting disturbance, the splitting of

the factor referred to in the student population into two factors in the psychiatric one may reflect the fact that some of the patients possess ego strength in seeing things differently (O , S) than other less adjusted patients. In the absence of known personality correlates of the patients' protocols, identification of the factors from a clinically meaningful point of view is difficult.

Hughes (1950) investigated the protocols of 100 psychiatric patients (32 organics, 39 psychoneurotics, 29 schizophrenics). Tetrachoric intercorrelations of 22 different signs representing the above syndromes were factored and yielded 8 orthogonal factors. The results were disappointing in that only the "organic" factor was clearly defined.

The factoring of persons (Q technique) on the basis of Rorschach scores was undertaken by Bendig and Hamlin (1955). Sixteen patients from four psychiatric categories were intercorrelated on 42 Rorschach element scores. The inverted factor analytic method (Q technique) did not separate the psychiatric categories although clinicians were able to do so via blind analyses. Apparently there is a great deal of meaningful information in the Rorschach protocol which is not represented in the scores obtained from it.

Factorial Relationship of Rorschach and MMPI

The relationship between the Rorschach and MMPI was investigated by Williams and Lawrence (1954) in a study which also included the Wechsler Scale of Adult Intelligence. The usual Rorschach "productivity" factor manifested no substantial MMPI subscale loadings. The second factor labeled "ego strength" was represented by all of the tests: Rorschach variables W (.82), k (.60), K (.59), C (.55), CF (.72), Wechsler Verbal IQ (.54), and MMPI variables K (.51), F (-.47), Sc (-.45) and Ego Strength (.71). A "repression" factor was not clearly defined but had highest loadings on the MMPI for the R (repression) scale (.55), D (.40), Hy (.54) and on the Rorschach by D (-.48), FM (-.49), FC (-.48), and R (-.44). The last factor was termed a "maladjustment" one and was heavily loaded with most of the MMPI subscales. Surprisingly, no

Rorschach variable was represented to any degree on this factor. The authors concluded that a considerable amount of variance in each test was not accounted for by covariation between them.

Rorschach as a Measure of Basic Personality

A study by Borgatta and Eschenbach (1955) utilized a wide range of Rorschach variables, aptitude scores, actuarial background data and personality ratings obtained from 125 normal male Air Force personnel. Three of the eight factors obtained loaded highly with Rorschach variables. These included a "productivity" factor, a "formal properties" factor, loaded highly on *C*, texture and shading, *S*, and *W*, which would be perhaps more accurately labeled as "low-perceptual form," and an "intelligence" factor. Of chief interest to the authors was the fact that the Rorschach did not contribute significantly to factors which they labeled as "social intelligence," "emotional assertiveness," "task leadership," "social acceptability" and "maturity adjustment." In view of these findings the authors believed that the use of the Rorschach as a tool for tapping the basic personality seemed questionable.

One or two kinds of M

A dimension of personality associating fantasy tendencies with the control of motility and impulsive behavior was hypothesized by Singer, Wilensky, and McCraven (1956). The population consisted of 100 male, veteran patients bearing the diagnosis of schizophrenia. These *Ss* received the Rorschach, TAT, and a battery of motor tests. Behavior rating scales were obtained from ward personnel. The *M* determinant was believed to represent both a tendency to inhibit motor activity and a considerable imagination or use of fantasy. While *M* loaded highly on two of the factors obtained, "motor inhibition" and "introspection," the fact that this hypothesized unitary factor emerged as two separate factors precluded a simple conclusion as to the meaning of *M*. The authors implied that the category *M* may be too gross in linking introversive individuals with those possessing considerable ability for re-

straining impulsive behavior. A key problem that remained unanswered was whether the factors might be a function of the population employed with little generality for more representative groups.

The Problem of "External" Validation

The difficulty of interpreting factors in clinical studies when no behavioral correlates are used is illustrated in a study by Vernier, Stafford, and Krugman (1958) who investigated the relationships among several projective tests (Rorschach, Sentence-Completion, Bender-Gestalt, and Draw-A-Person). Another purpose was to determine the relationships between four kinds of physical pathology (orthopedic, respiratory, cardiac, and neurological) and the projective scores. The *Ss*, 150 VA patients (105 in-patients, 45 out-patients) received the Otis S. A. Test of Mental Ability in addition to the aforementioned projective techniques.

Tetrachoric correlations of 53 variables obtained from the tests were reduced to a 35×35 matrix by eliminating intercorrelations approaching zero as well as those over .75. The analysis yielded seven oblique factors which are extremely unclear in meaning. The lack of clarity is a function of the lack of experimental evidence as to the import of many of the "signs" obtained from projective techniques. For example, the "withdrawal" factor had substantial loadings on "hands omitted" from the Draw-A-Person test (.63) and on the "need affiliation" score from the Sentence Completion test (-.49). What this has to do with a third highly loaded variable "not rejecting cards on the Rorschach" (.64) is unknown. Do people who withdraw from interpersonal contact tend not to reject Rorschach cards? Similarly, the six other factors obtained are difficult to interpret. The weakness of this study lies in the fact that the interpretive implications of clusters of signs from projective techniques are themselves meaningless unless we know the personological correlates of these responses. The crucial task then for researchers employing factor analysis is to indicate the common variance between projective responses and behavior in important personal-social situations, rather than indicating the common variance between

responses from different projective techniques.

A study which deals with this problem is that of Sandler and Ackner (1951). Here the "external" validation was obtained by correlating 200 odd psychiatric ratings with the Rorschach factors treated as scores using the formula given by Sandler (1949). Space does not permit a full discussion of these interesting data. Suffice it to mention that Factor 2, for example (the perception of internal, anatomical objects as opposed to the perception of external objects), correlated beyond the .01 level with such current symptoms of the patient as "overt aggressive behavior" (.53), "complaints of feelings of aggression" (.41), "confabulations" (.41) as well as with previous personality traits of "appearing self-distrustful" (.42) and "feeling tense" (.36). This interesting study shows the possibilities of obtaining new information when sufficiently broad external behavior scores are compared with factors treated as scores.

PRESENT INNOVATIONS IN THE RORSCHACH

In the foregoing pages an attempt has been made to discuss the problems attendant to the use of the Rorschach as a psychometric instrument and at the same time to present some of the important findings resulting from factorial studies. In a way the two tasks are contradictory. If the Rorschach is so ill-fitted for psychometric treatment why bother to review studies at all? The reason for reviewing the content of many of these studies is that they present an admixture of "true" and "error" variance which is difficult to separate so long as we retain the Rorschach in its present form. In an effort to retain the baby we have saved the dirty bath water, too. Though the Rorschach may be a means of probing the depths of personality in the hands of some practicing clinicians, from the point of research it must be considered a psychometric sow's ear. The small number of cards, quasiquantitative scoring system with resulting low reliability, variable number of *R*, use of multiple scores from single responses, attests to its ancestry. Yet, another difficulty lies in the fact that the scoring is completely dependent on the self-knowledge, verbal fluency, and vocabulary strength of the testee, not to mention the skill of the

examiner in asking the right questions. The Rorschach is operationally then hardly a *perceptual test* but rather an *interpretation of an interpretation*. Several researchers have attempted to remedy this situation. Cronbach (1949) has offered several statistical approaches to the Rorschach. He favors counting procedures over additive methods. Where correlation is desired the scores should be normalized. Ratio and difference scores should be replaced by patterns. His suggestions have had some influence on Rorschach research. They are, however, of a "let's make the best of a bad situation" kind and are intended as stopgap measures pending the arrival of a more statistically sophisticated ink blot technique.

Rust (1947), McReynolds (1951), and O'Reilly (1956) have constructed interesting subsystems designed for specialized use either in determining the presence of a given determinant, or for psychiatric diagnosis. In the interest of space, however, only those systems dealing with a new total scoring system will be considered. The Howard Ink Blot test (Howard, 1953) does contain 12 new blots, but the system of scoring employed is identical to the Beck system and there is therefore little point in reviewing the technique.

Zubin (1953), the first to apply more adequate quantitative methodology, constructed 60 scales measuring location, objective stimulus attributes, determinants, interpretation, content, organization and other attributes somewhat difficult to categorize. The scores obtained are rated on a five-point scale. Zubin's system has not been generally accepted for two reasons. One, a very basic reason, is that the system has not until very recently been published in book form. Secondly, the excessive number of characteristics is a bit too detailed for the current state of knowledge about the stimulus correlates of the Rorschach.

A more recent variation is that of Holtzman (1959) who employs his own set of ink blots. He has borrowed Zubin's rating scheme, trimming the number of scoring categories utilized, however, to six. These are Location, Form Appropriateness, Form Definiteness, Color, Shading, and Movement En-

ergy Level, which are rated on from three- to seven-point scales. In addition, several new innovations are introduced for the purpose of making the scores more amenable to treatment as normal continuous data. Two sets of parallel cards are employed, each set containing 45 cards. With this large set of cards it was possible to ensure that fact that not only would each card be used, but that the same number of responses would be given to each card. Accordingly, the *S* is instructed to give only one response to each card.

The system is a major advance in obtaining higher reliabilities in the scoring of the categories. Interscorer reliability is in the .90's, while estimates of reliability based on internal consistency using Guilliksen's matched random subtest method yielded correlations ranging from .80 to .91.

In view of these important advances it seems unfortunate that in one respect the system has remained stagnant. This is in the manner of recording the response. Here Holtzman has relied more or less on the highly unsatisfactory standard inquiry method. The disadvantage of this method, which relies upon the *S*'s ability to verbalize the determinants of his perception, is apparent in the data presented by Baughman (1954). It is known that most people who see a bat on Card I give form as the chief determinant. Very few mention the black color and shading as influencing their percept. When, however, Baughman presented a modified version of Card I which had only the color removed, very few people saw the bat. Further evidence in this regard is reported in a recent study comparing two methods of inquiry (Baughman 1959). One method was the standard Beck approach. The other was the paired-comparison inquiry. This method, briefly stated, presents the orthodox card and one of several various modifications during the inquiry period. The modifications (Baughman, 1958) currently employed are as follows:

Achromatic (A): Color is eliminated; remaining blot properties are retained as in the standard series. Complex Silhouette (CS): Shading variations are removed; differentiation between major detail areas is retained, however, by giving each major detail area a different contrast value. Silhouette (S): All

shading variations and differentiations between major detail areas are eliminated, making a uniform gray. White (W): A white figure is placed on a uniformly gray background. Complex Form (CF): Figure-ground contrast due to brightness difference is removed; forms of major detail areas are retained. The form is like that of CS. Form (F): A form comparable to S and W is created, but figure-ground contrast due to brightness difference is eliminated (p. 382).

The examiner chooses the modification to suit his question. If, for example, *S* sees "bat" on Card II and *E* wishes to determine the role of shading in the perception, the silhouette Card II might be compared to the standard chromatic one, and *S* asked if he still sees the bat.

The Beck-Baughman inquiry comparison indicated that much more shading was actually utilized by most *Ss* on the Rorschach than was verbalized under the Beck inquiry.

The Baughman method which retains the original 10 cards will probably not supply quite as reliable data as the more refined statistical procedures seem to give the Holtzman modification. On the other hand Baughman's studies have the advantage of being more readily related to the vast Rorschach literature and should prove of greater interest to most practicing clinicians. Both have the disadvantage of requiring an excessive number of cards, 45 in each of the two Holtzman series, and 65 in Baughman's group. It is too early to state which of these procedures will prove to be the most fruitful for personality study but both represent a considerable improvement over the current Rorschach procedure.

The importance of these new revisions for factor analysis is considerable. To the extent that the current Rorschach is inadequate as a quantitative test the use of factor analysis will yield questionable results. If factor analysis is to be utilized more fully in the future, the Rorschach will have to be superseded by a more appropriate test. The newer variations with increasing reliance on quantification and reliability should more readily justify the use of factor analyses and increase our confidence in the results obtained from such studies.

REFERENCES

- ADCOCK, C. J. A factorial approach to Rorschach interpretation. *J. gen. Psychol.*, 1951, 44, 261-272.

- AINSWORTH, MARY D. Problems of validation. In Klopfer, B. and Others, *Developments in the Rorschach technique*. New York: World Book, 1954, Pp. 405-500.
- BAUGHMAN, E. E. A comparative analysis of Rorschach forms with altered stimulus characteristics. *J. proj. Tech.*, 1954, 18, 151-164.
- BAUGHMAN, E. E. A new method of Rorschach inquiry. *J. proj. Tech.*, 1958, 23, 381-389.
- BAUGHMAN, E. E. The effect of inquiry method on Rorschach color and shading scores. *J. proj. Tech.*, 1959, 23, 3-7.
- BENDIG, A. W., & HAMLIN, R. M. The psychiatric validity of an inverted factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1955, 19, 183-188.
- BORGATTA, E. F., & ESCHENBACH, A. E. Factor analysis of Rorschach variables and behavioral observation. *Psychol. Rep.*, 1955, 1, 129-136.
- COAN, R. A factor analysis of Rorschach determinants. *J. proj. Tech.*, 1956, 20, 280-287.
- CONSALVI, C., & CENTER, A. Rorschach scores as a function of four factors. *J. consult. Psychol.*, 1957, 21, 47-51.
- COX, S. M. A factorial study of the Rorschach responses of normal and maladjusted boys. *J. genet. Psychol.*, 1951, 79, 95-115.
- CRONBACH, L. J. "Pattern Tabulation": A statistical method for analysis of limited patterns of scores, with particular reference to the Rorschach test. *Educ. Psychol. Measmt.*, 1949, 9, 149-171.
- FISKE, D. W., & BAUGHMAN, E. E. Relationships between Rorschach scoring categories and the total number of responses. *J. abnorm. soc. Psychol.*, 1953, 48, 25-32.
- FOSTER, A. The factorial structure of the Rorschach test. *Tex. Rep. Biol. Med.*, 1955, 13, 34-61.
- GLICKSTEIN, M. A note on Wittenborn's factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1959, 23, 69-75.
- HOLTZMAN, W. H. Objective scoring of projective tests. In B. M. Bass & I. A. Berg (Eds.), *Objective approaches to personality assessment*. New York: Van Nostrand, 1959, Pp. 119-145.
- HOWARD, J. W. The Howard Ink Blot Test: A descriptive manual. *J. clin. Psychol.*, 1953, 9, 209-254.
- HSÜ, E. H. The Rorschach responses and factor analysis. *J. gen. Psychol.*, 1947, 37, 129-138.
- HUGHES, R. M. A factor analysis of Rorschach diagnostic signs. *J. gen. Psychol.*, 1950, 43, 85-103.
- LEVIN, M. The two tests in the Rorschach. *J. proj. Tech.*, 1953, 17, 471-473.
- LOTSOF, E. J. Intelligence, verbal fluency, and the Rorschach test. *J. consult. Psychol.*, 1953, 17, 21-24.
- LOTSOF, E. J., COMREY, A., BORGARTZ, W., & ARSFIELD, P. A factor analysis of the WISC and Rorschach. *J. proj. Tech.*, 1958, 22, 297-301.
- MCREYNOLDS, P. Perception of Rorschach concepts as related to personality deviations. *J. abnorm. soc. Psychol.*, 1951, 46, 131-141.
- MURSTEIN, B. I. Review of Klopfer, B. & others. *Developments in the Rorschach techniques. Vol. II. Fields of application. J. proj. Tech.*, 1958, 22, 246-248.
- O'REILLY, B. O. The objective Rorschach: A suggested modification of Rorschach techniques. *J. clin. Psychol.*, 1956, 12, 27-31.
- RUST, R. M. Some correlates of the movement response. *J. Pers.*, 1947, 16, 369-401.
- SANDLER, J. The reciprocity principle as an aid to factor analysis. *Brit. J. Psychol. stat. sect.*, 1949, 2, 180-187.
- SANDLER, J., & ACKNER, B. Rorschach content analysis: an experimental investigation. *Brit. J. med. Psychol.*, 1951, 23, 180-201.
- SEN, A. A study of the Rorschach test. *Brit. J. Psychol., statist. sect.*, 1950, 3, 21-39.
- SINGER, J. L., WILENSKY, H., & MCCRAVEN, V. G. Delaying capacity, fantasy, and planning ability: A factorial study of some basic ego functions. *J. consult. Psychol.*, 1956, 20, 375-383.
- STOTSKY, B. A. Factor analysis of Rorschach scores of schizophrenics. *J. clin. Psychol.*, 1957, 13, 275-278.
- VERNIER, CLAIRE M., STAFFORD, J. W., & KRUGMAN, A. D. A factor analysis of indices from four projective techniques associated with four different types of physical pathology. *J. consult. Psychol.*, 1958, 22, 433-439.
- WEBB, W., & HILDEN, A. H. Verbal and intellectual ability as factors in projective test results. *J. proj. Tech.*, 1953, 17, 102-103.
- WILLIAMS, H. L., & LAWRENCE, J. F. Further investigation of Rorschach determinants subjected to factor analysis. *J. consult. Psychol.*, 1953, 17, 261-264.
- WILLIAMS, H. L., & LAWRENCE, J. F. Comparison of the Rorschach and MMPI by means of factor analysis. *J. consult. Psychol.*, 1954, 18, 193-197.
- WITTENBORN, J. R. A factor analysis of discrete responses to the Rorschach ink blots. *J. consult. Psychol.*, 1949, 13, 335-340.
- WITTENBORN, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, 14, 261-267. (a)
- WITTENBORN, J. R. Level of mental health as a factor in the implications of Rorschach scores. *J. consult. Psychol.*, 1950, 14, 469-472. (b)
- WITTENBORN, J. R. Some comments on confounded correlations among Rorschach scores. *J. consult. Psychol.*, 1959, 23, 75-77.
- ZUBIN, J., & ERON, L. *Experimental abnormal psychology* (Prelim. ed.). New York: New York State Psychiat. Inst., 1953.

(Received July 13, 1959)

BRIEF REPORTS

RELATION OF RORSCHACH MOVEMENT AND COLOR RESPONSES TO COGNITIVE INHIBITION^{1, 2}

ROBERT E. FAGER

State University of New York³

Recent research has suggested that Rorschach human movement (*M*) may be considered both as a product and as a measure of inhibition ability. Levine, Glass, & Meltzoff (1957) proposed that failure to make a mirror-image *N* on the digit symbol subtest of the Wechsler-Bellevue Intelligence Scale is a function of insufficient delay of the overlearned *N* response tendency. They hypothesized that the *Ss* who fail to reverse the *N* (experimentals) should produce fewer *M* responses than *Ss* who make the reversal (controls). They found: (a) greater proportion of experimentals than controls produced less than two *M* responses, (b) experimentals more frequently than controls had *Sum C* = 0, (c) experimentals had an experience balance characterized by Low *M*-Low *Sum C* group, while controls usually fell into the High *M*-High *Sum C* group. The purpose of the study reported here is to replicate Levine's work.

Whereas Levine used only male psychiatric veterans, this study employed male and female *Ss* from both psychiatric and nonpsychiatric hospital wards. There were no differences in sex distribution between experimental and control groups, although they did differ in age and education (controls older and better educated). Psychiatric and nonpsychiatric *Ss* were combined because no significant differences were found for intelligence, total Rorschach responses, age or education. The IQ-*M* correlation for the total sample (+.22, $p < .01$) was of the order reported by Levine et al. (1959) for other studies (+.26). Ninety-nine of the 124 controls and 58 of the 81 experimentals were psychiatric patients representing a wide variety of disorders. All *Ss*

were 16 years or older and made 10 or more Rorschach responses. Experimentals were identified by noting errors in copying the mirror-image *N*; and controls were selected by taking the next protocol in alphabetical order which did not have this error.

Results show no group differences (a) in production of less than two *M* responses; (b) in proportion of *Sum C* = 0; and (c) in type of experience balance. Moreover, differences in neither mean IQ (experimentals 97.9, controls 101.9) nor mean total responses (experimentals 23.1, controls 24.8) reached significance.

Since Levine had significant IQ differences in favor of controls ($p = .0001$), an effort was made to replicate this feature by comparing high IQ controls (mean 117.7) with low IQ experimentals (mean 80.0). Again, neither the proportion making less than two *M* responses nor the frequency of *Sum C* = 0 discriminated between groups.

Failure to replicate Levine's results raises the question of the stability of his findings. Sampling differences are an unlikely explanation for the variant findings since no demonstrable bias was introduced by including females and nonpsychiatric *Ss*. An interaction between intelligence and inhibition is also a doubtful explanation since differences in *M* and *Sum C* could not be demonstrated in this study even when alternate experimental and control groups were selected to maximize intellectual differences. Consequently, the relation of Rorschach human movement and color responses to the type of cognitive inhibition employed by Levine remains unclear.

REFERENCES

- LEVINE, M., GLASS, H., & MELTZOFF, J. The inhibition process, Rorschach human movement responses, and intelligence. *J. consult. Psychol.*, 1957, 21, 41-45.
- LEVINE, M., SPIVACK, G., & WIGHT, B. The inhibition process, Rorschach human movement responses, and intelligence. *J. consult. Psychol.*, 1959, 23, 306-311.

(Received June 5, 1959)

¹ An extended report of this study may be obtained without charge from Robert E. Fager, Department of Psychiatry, Upstate Medical Center, Syracuse 10, New York, or for a fee from the American Documentation Institute. Order Document No. 6288, remitting \$1.25 for microfilm or \$1.25 for photocopies.

² These data were collected while the author was at the Department of Psychiatry, State University of Iowa.

³ Department of Psychiatry, Upstate Medical Center.

A FURTHER INVESTIGATION OF BODY-CATHEXIS AND THE SELF^{1, 2}JON R. WEINBERG³*State University of Iowa*

The present study essentially replicated an experiment reported by Secord and Jourard (1953), in which attitudes toward one's body were related to other aspects of the self. The relationships between a number of pairs of variables were again investigated by correlational analysis. The following measures were employed: the homonym test, a word association procedure from which a score on somatic concern is derived; the body-cathexis scale, composed of self-ratings of satisfaction with parts of the body; the anxiety indicator, a subscore consisting of body-cathexis items given the lowest average ratings; the self-cathexis scale composed of self-ratings of satisfaction with personal attributes; and the Maslow test of psychological insecurity, a questionnaire instrument.

The Ss, students in an introductory psychology course, were tested in groups of 30 to 40. Conditions of complete anonymity were insured, and honesty in responding was strongly emphasized in the instructions. The scores of 108 men and 104 women were employed in the analysis.

¹ An extended report of this study may be obtained without charge from Jon Weinberg, C 605 Mayo Hospital, Minneapolis 14, Minnesota, or for a fee from the American Documentation Institute. Order Document No. 6289, remitting \$1.25 for microfilm or \$1.25 for photocopies.

² This paper is based upon a master's thesis submitted to the State University of Iowa. The author is grateful to A. L. Benton for his guidance in this research.

³ Now a United States Public Health Service Postdoctoral Fellow at the University of Minnesota Medical School.

The primary interest of the study was reassessment of the correlations between pairs of measures. The results indicated that some of the significant relationships found previously were replicable, while others were not. Briefly, it was again found that body-cathexis and self-cathexis were significantly related; but in the present study, the correlation for males was significantly higher than that for females. In contrast to the Secord-Jourard findings, the homonym test was not significantly related to the anxiety indicator nor, for women, to body-cathexis. The Maslow insecurity test was again significantly related to both body-cathexis and self-cathexis, but to the anxiety indicator for men only. A significant correlation for men was also found between the Maslow test and the homonym test, a relationship not investigated earlier.

In general, then, it was again found that feelings about the body are related both to feelings about the self and to feelings of psychological insecurity. Both of these relationships, however, tended to be of greater magnitude for men than for women in the present study. The finding by Secord and Jourard that negative feelings about the body are related to somatic concern was not upheld. In addition, the present study found that, for men, somatic concern was related to psychological insecurity.

REFERENCE

- SECORD, P. F., & JOURARD, S. M. The appraisal of body-cathexis: Body-cathexis and the self. *J. consult. Psychol.*, 1953, 17, 343-347.

(Received August 27, 1959)

THE TEMPORAL STABILITY OF MPI SCORES IN NORMAL AND PSYCHIATRIC POPULATIONS¹

JOHN B. KNOWLES

Medical Research Council, Clinical Psychiatry Research Group, Chichester, Sussex

The Maudsley Personality Inventory (MPI) constructed and standardized in Britain by Eysenck (1959) gives measures of Neuroticism (N) and Extraversion (E) as defined within his personality theory. The present study was designed to assess the test-retest reliability of the scale and, by comparing the retest correlations for psychiatric and normal Ss, to test hypotheses relating to its validity.

Eysenck postulates that the relative predominance of either cortical excitation or inhibition predisposes individuals towards introverted or extraverted patterns of behavior. If differences in excitation and inhibition are properties of the nervous system, then measures of personality based on these differences should remain relatively stable over time (Storms & Sigal, 1958). Further, it was predicted that normal and psychiatric populations would show an equivalent degree of stability over time, as measured by the test-retest correlation. Since neuroticism is defined as a liability to breakdown under stress, neither the appearance of, nor fluctuation in clinical symptoms, should affect the temporal stability of N scores. Thus, it was predicted that the test-retest correlations for N in neurotic and normal populations would not differ significantly.

Ninety-three surgical patients, and 63 psychiatric patients with a clearly neurotic diagnosis, who had completed the MPI one year previously, were contacted through the post and asked to complete a further MPI. Fifty-two surgical patients, used as the normal group in subsequent comparisons, and 40 neurotic patients replied.

¹ An extended report of this study may be obtained without charge from John B. Knowles, Medical Research Council, Clinical Psychiatry Research Group, Graylingwell Hospital, Chichester, Sussex, or for a fee from the American Documentation Institute. Order Document No. 6291, remitting \$1.25 for microfilm or \$1.25 for photocopies.

The retest correlations for E were 0.77 and 0.79 for the normal and neurotic patients respectively. Since these values were almost identical, and virtually the same as the split-half reliability of the scale (approximately 0.80) it was concluded that the predictions for E were confirmed.

The results appeared to disfavor the prediction regarding the N factor. The retest correlation for the neurotics (0.70) was below that for the normals (0.84). However, when tested by Fisher's *z* transformation the difference just failed to reach significance at the 5% level.

The present results agree with those given for normals by Eysenck, and for two mixed normal and neurotic groups by Bartholomew & Marley (1959). When the groups studied in the latter investigation are compared, it is found that the group which was predominantly neurotic also gave a lower retest correlation on neuroticism whereas the correlations for E were similar.

From the studies undertaken to date it is concluded that the MPI has high reliability, that E scores of normal and neurotic populations show an equivalent stability over time, but that there is some evidence that fluctuations in clinical status influence the stability of N scores.

REFERENCES

- BARTHOLOMEW, A. A. & MARLEY, E. The temporal reliability of the Maudsley Personality Inventory. *J. ment. Sci.*, 1959, 105, 238-240.
- EYSENCK, H. J. *Manual of the Maudsley Personality Inventory*. London: Univer. of London Press, 1959.
- STORMS, L. M., & SIGAL, J. J. Eysenck's personality theory with special reference to "The dynamics of anxiety and hysteria." *Brit. J. med. Psychol.*, 1958, 31, 228-246.

(Received August 27, 1959)

DIMENSIONS OF SOCIAL DESIRABILITY¹

SAMUEL MESSICK

Educational Testing Service, Princeton, New Jersey

A high correlation has been repeatedly noted between the judged desirability of a personality characteristic denoted by an inventory item and the probability of the item's endorsement (cf. Edwards, 1957b; Wahler, 1958). Edwards (1953) has pointed out that, while this correlation suggests the possibility that traits judged to be desirable are prevalent or dominant in the group under study, it might also reflect a tendency to create a good impression or to respond, either deliberately or unintentionally, in a socially desirable manner. Instruments designed to measure this set to respond desirably were found to be substantially correlated with several clinical and personality scales (Edwards, 1954; Edwards, 1957b; Fordyce, 1956; Hanley, 1957) thus documenting the pervasive influence of social desirability on personality measurement. In view of this accumulating evidence, Edwards (1954) and others (cf. Heineman, 1953; Sisson, 1948) attempted to control for social desirability in constructing personality measures by pairing, in a forced-choice format, items with approximately equal desirability scale values.

This forced-choice method of control assumes that social desirability is a single unidimensional continuum, along which statements may be spaced and in terms of which items may be equated. If, however, judgments

of social desirability were found to be multidimensional, then controls in terms of a single average scale might prove inadequate. Edwards (1953) obtained his social desirability scale values by the method of successive intervals (cf. Edwards, 1957a; Torgerson, 1958), a technique which combines individual ratings of desirability to produce a single scale describing, on the average, the pooled judgments of the particular group used. Since the subjects' (Ss') own opinions about the social desirability of traits were elicited in the rating task, one might expect various cultural and social background factors to produce differences in the scales obtained from different judging groups. However, there may also exist within a single group consistent individual differences with sufficient structure to warrant their consideration as different "points of view" with respect to desirability. The purpose of the present paper is to investigate the dimensionality of social desirability judgments and to explore the nature of any dimensions or "points of view" uncovered.

Edwards (1957b) and others (cf. Klett & Yaukey, 1959) have investigated the comparability of social desirability judgments obtained from different groups and have found generally high correlations among the various scales. On a college sample, for example, Edwards (1953) found no essential differences in desirability values as a function of age, sex, or education, and Klett (1957a), on a high school sample, concluded that there were no significant differences among socioeconomic groups, or between grades or sexes. The correlation between desirability scale values obtained from the college and high school samples was .94. Klett (1957b) also obtained desirability ratings from a sample of mental hospital patients; he found no essential differences in the desirability ratings

¹ This project was supported by the Office of Naval Research, under Contract Nonr-2214(00). The study was greatly facilitated by C. James Klett, who graciously supplied his raw data for reanalysis. The author also wishes to thank Ledyard Tucker for his helpful advice on the analysis and rotations, Henrietta Gallagher for supervising the computations, Henry Kaiser for arranging to have an oblique rotation performed on the Iliac computer, and Douglas N. Jackson for his perceptive comments on the manuscript. Portions of this paper were read at the American Psychological Association Meeting in Cincinnati, 1959.

of psychotic and nonpsychotic subgroups, and scale values for his total hospital sample correlated .88 with the college values and .87 with the high school values. In evaluating this striking similarity in group opinions about the desirability of personality traits, it must be remembered that successive intervals scale values represent a kind of average rating for the group involved. Within-group ratings would need to be very homogeneous and between-group differences quite marked before a low correlation would be expected between such averages. Rosen (1956), for example, reported highly similar mean profiles for groups given the MMPI under self-appraisal and desirability instructions, but at the individual level the correlations between normal and desirability profiles were low and sometimes negative.

Despite high correlations between scale values obtained from different groups, evidence suggesting the inadequacy of a single global view of desirability has been noted. Klett (1957a) found that matched item pairs for the Edwards Personal Preference Schedule (EPPS) were less adequately matched in terms of the high school values. French (1958) noted that EPPS scores for basic airmen were significantly higher than published norms on traits presumably valued by the Air Force, suggesting that recruits were able to identify responses that get military approval and that the Air Force view of desirability may differ from that expressed in the college ratings. Klett (1957b), noting that some of the differences between hospital and college ratings were systematically related to scale content, found that EPPS mean scores for a sample of schizophrenics were significantly higher than college normative scores for needs judged to be more desirable by the hospital group. Klett (1957c) also divided the mental hospital sample into several dichotomous groups in terms of high and low scores on scales derived from the MMPI. For each breakdown, correlations were computed between median desirability ratings obtained from high and low scorers; and some of the coefficients, although still high (.68 and .71), were significantly smaller than those previously reported.

The present study carries the investigation

of consistent differences in judging desirability back to the level of the individual ratings. The technique applied is essentially a factor analysis of intercorrelations among ratings of personality items. This approach, used previously by Morris and Jones (1955) in their study of value dimensions, yields factors which presumably reflect different points of view with respect to the attribute rated, in this case social desirability.

An alternative approach to this same problem would be to analyze "*Q* correlations" between people to isolate groups of similar raters and examine their median patterns. This use of *Q* correlations, however, depends upon the existence of relatively pure types of raters; whereas for the present approach, which seeks consistent judgmental tendencies, it is an empirical question whether or not such types exist, i.e., whether there is a simple structure among factor scores for the raters. In the present analysis dimensions of desirability can emerge even though raters are "complex," i.e., have high factor scores on more than one dimension, which is the usual case in the area of mental abilities. In this question of research strategy, it was considered more fruitful to deal with a structured sample of items than to depend upon the isolation of a typology of raters. In any event, factor spaces obtained from the two approaches are closely related, and the distribution of people in the space may be examined in the present method by the subsequent determination of factor scores.

METHOD

Subjects

The present study is a reanalysis of desirability judgments collected by Klett (1957b) from 118 manifestly disturbed mental hospital patients. For the present analysis 10 Ss were eliminated from the sample because of incomplete data.

Procedure and Analysis

The Ss had been asked to rate on a nine-category scale the desirability of 140 items purportedly representing 14 psychological needs. Categories ran from one for "extremely undesirable" through five for "neutral" to nine for "extremely desirable." These items had been previously scaled by Edwards (1954) in constructing measures for the EPPS.

In Klett's (1957b) analysis of this material, the scaling procedure of successive intervals (cf. Torger-

TABLE 1
PERSONALITY ITEMS SELECTED FOR
FACTOR ANALYSIS

Item	EPPS Item No.	Scale	Median Rating
1	17A	Deference	7.13
2	178A	Endurance	8.28
3	114A	Dominance	5.97
4	2A	Deference	8.00
5	68A	Order	8.21
6	152A	Change	8.20
7	41A	Achievement	5.97
8	45A	Autonomy	4.10
9	14A	Exhibition	6.80
10	78A	Succorance	7.82
11	221A	Nurturance	6.74
12	5A	Autonomy	8.41
13	214A	Heterosexuality	7.97
14	11A	Achievement	7.56
15	222A	Change	6.31
16	98A	Succorance	6.30
17	39A	Exhibition	5.44
18	82A	Intracception	6.47
19	94A	Dominance	4.55
20	224A	Heterosexuality	5.60
21	168A	Endurance	7.07
22	42A	Deference	6.09
23	91A	Affiliation	7.49
24	169A	Heterosexuality	5.78
25	146A	Affiliation	6.81
26	33A	Order	7.32
27	51A	Achievement	8.40
28	190A	Aggression	5.24
29	102A	Intracception	7.64
30	180A	Aggression	6.32
31	70A	Autonomy	5.67
32	65B	Endurance	7.09
33	93A	Succorance	5.36
34	76A	Affiliation	8.24
35	49A	Exhibition	4.47
36	206A	Nurturance	8.08
37	185A	Aggression	3.28
38	18A	Order	7.21
39	166A	Nurturance	7.42
40	104A	Dominance	7.34
41	197A	Change	6.60
42	112A	Intracception	6.17

son, 1958) was used, which in effect simultaneously normalized the ratings of the 140 items on a common base line. The scale values of the items were essentially the means of these normalized ratings as measured on the common scale. In the present analysis, on the other hand, the original ratings of each item were intercorrelated and factor analyzed. In a sense the two analyses complement each other in that the first one dealt with means and variances to obtain an average scale for the group, and the second dealt with intercorrelations to obtain the

structure of consistent individual differences in judgments.

To make the factor analysis of feasible size, 42 items were selected from the 140 available, three from each of the 14 needs represented. By examining the distribution of ratings for each statement, one item was chosen with a high median desirability value for each need, one with an intermediate value, and one with a low value. No attempt was made to select in terms of item content, except to require three items for each need. Since the needs differed in their judged desirability, the median values for items representing one need overlapped with those of another to cover a wide range of desirability. This variation in judged desirability within each need favored the appearance of a general desirability factor, but at the same time the choice of three items from each scale permitted the occurrence of clustering according to needs. EPPS item numbers and median desirability ratings are presented in Table 1 for the statements selected. Product-moment correlations among the ratings of these 42 items were factor analyzed by the grouping method (Thurstone, 1947), with communalities estimated by the largest correlation in each column.

RESULTS

After the extraction of nine factors, the residuals were found to be symmetrically distributed about zero with 93% of the coefficients included between .10 and -.10. Since further factor extraction would have generated small doublets of doubtful interpretation, only nine factors were retained for rotation. Although one of them (Factor IX) was quite large, accounting for approximately 42% of the common variance and 22.6% of the total variance, a general factor of desirability did not seem indicated. For comparison purposes, the first centroid factor was also extracted; it accounted for 22.8% of the total variance. The factors were rotated analytically to oblique simple structure using the oblimax criterion (Pinzka & Saunders, 1954), and some minor additional rotations were performed graphically. The final rotated loadings are given in Table 2. In the orthogonal solution obtained by the grouping method, the nine factors accounted for the following percentages of total variance, respectively: 4.03, 3.57, 3.39, 9.05, 3.19, 2.48, 2.57, 2.53, 22.65. These orthogonal factors were naturally modified, sometimes considerably, in the course of rotation, but they correspond roughly to the oblique factors in the order listed in Table 2. The intercorrelations among

TABLE 2
FACTOR LOADINGS ROTATED OBLIQUELY

Variable	I	II	III	IV	Factor V	VI	VII	VIII	IX
1	.22	.12	-.06	-.01	-.12	-.26	-.15	.35	.31
2	-.10	-.04	-.06	-.09	.21	.03	.01	.16	.39
3	.31	.06	.03	.19	.08	-.01	.03	.03	.10
4	.01	-.00	-.18	.11	-.06	.08	.19	.10	.40
5	-.23	-.00	-.00	.01	.02	.25	-.03	.14	.34
6	.05	-.00	.27	-.13	.03	-.18	.55	-.11	.51
7	.56	-.04	.18	-.09	.06	.07	-.01	-.05	.08
8	-.13	.03	.28	.04	.10	-.00	.08	.03	-.27
9	.04	-.04	.26	.04	.25	-.05	-.04	-.04	.20
10	-.04	.02	.14	-.11	.02	-.00	.01	.15	.31
11	.43	.04	.39	-.11	.04	.00	-.14	-.13	.03
12	-.19	-.04	.02	-.02	.09	.08	.18	-.01	.46
13	-.22	-.13	.19	.04	.11	.07	-.00	.08	.12
14	.23	-.17	.23	.02	.08	-.21	.10	.04	.36
15	.01	.06	.35	.07	-.03	.16	.18	.16	-.17
16	.10	.06	.67	.06	-.08	-.05	-.07	-.01	-.08
17	-.03	.25	.48	.07	.06	-.03	-.10	-.06	.12
18	.64	-.12	.04	.11	-.05	-.01	.11	.14	.02
19	.10	.04	.25	.22	.21	.02	-.13	-.02	-.19
20	-.03	.01	.06	.64	.00	.02	-.05	.05	-.11
21	-.21	.15	-.04	-.11	.13	.27	-.18	.10	.13
22	-.02	.24	-.09	.13	-.06	.05	-.01	.43	-.06
23	.03	.03	.14	.16	-.17	.31	.04	.03	.29
24	-.05	-.05	.10	.59	-.02	-.02	.10	-.04	-.01
25	-.00	-.01	-.00	-.11	-.01	.67	-.01	.03	-.11
26	-.01	-.02	-.02	-.12	.03	.02	-.01	.52	.02
27	-.12	-.01	-.02	-.04	-.03	.14	.10	.06	.56
28	-.20	.30	.13	.23	.07	.03	.05	-.16	.07
29	-.04	.08	.04	-.05	.12	.01	.25	-.05	.44
30	-.05	.09	-.01	-.04	.56	-.07	-.01	.05	.04
31	-.16	.06	-.07	.15	.44	.09	-.10	.01	-.08
32	.21	.23	-.22	.25	.00	-.06	-.16	-.14	.61
33	.06	.29	.51	.15	-.07	-.07	-.12	.01	-.01
34	-.05	-.05	.11	-.14	-.02	.20	.20	-.02	.48
35	.02	.58	.31	-.01	.05	-.05	.09	-.08	.07
36	.27	-.10	.11	.17	-.12	.03	-.11	.04	.45
37	-.16	.62	.08	-.03	.02	.11	.07	.07	-.10
38	.06	-.08	.08	.22	.10	.09	-.11	.35	-.03
39	-.13	-.04	.27	.08	-.01	.41	.07	-.13	.08
40	.05	-.14	-.00	-.02	.54	-.10	.10	-.10	.29
41	-.01	.02	-.04	.07	-.02	.10	.69	-.00	-.02
42	.03	.13	-.06	.01	.36	.19	-.04	-.00	.05

primary factors (Table 3) are generally positive and indicate the presence of a second-order structure, but enough coefficients are sufficiently low to suggest that a general factor of desirability does not emerge at the second-order level either.

These factors summarize certain consistencies among the original ratings and may be interpreted in terms of different viewpoints as to what is desirable. Factor loadings .25 or greater will be emphasized in the inter-

pretations. In general the items do not cluster according to the specific scales they were written to represent, but it should be noted that there is no necessary relationship between these factors of *judged* desirability and the scales or item clusters determined from *responses* to the inventory. The fact that the items do not cluster according to specific needs when rated for desirability does not reflect upon the homogeneity of scales from Edwards' inventory.

TABLE 3
INTERCORRELATIONS AMONG PRIMARY FACTORS

Factor	I	II	III	IV	V	VI	VII	VIII	IX
I	1.00	.11	-.03	.11	.15	.19	.16	.17	.03
II	.11	1.00	-.05	.16	.10	-.18	-.03	-.21	-.36
III	-.03	-.05	1.00	.35	.22	.28	.08	.48	.30
IV	.11	.16	.35	1.00	.49	.20	.33	.24	.12
V	.15	.10	.22	.49	1.00	.29	.27	.24	.29
VI	.19	-.18	.28	.20	.29	1.00	.17	.49	.58
VII	.16	-.03	.08	.33	.27	.17	1.00	.30	.15
VIII	.17	-.21	.48	.24	.24	.49	.30	1.00	.48
IX	.03	-.36	.30	.12	.29	.58	.15	.48	1.00

Factor I

18. To like to put yourself in someone else's place and to imagine how you would feel in the same situation. (Int.) .64
7. To like to write a great novel or play. (Ach.) .56
11. To like to have your friends confide in you and tell you their troubles. (Nur.) .43
3. To like to settle arguments and disputes between other people. (Dom.) .31
36. To like to treat other people with kindness and sympathy. (Nur.) .27

The common behavioral element viewed as desirable on this particular dimension might be labeled Interpersonal Concern or Interpersonal Sensitivity. The determination of factor scores would presumably uncover individuals in the sample analyzed with consistent tendencies to rate these items highly desirable, even though the scale values for the total group are only moderate (see Table 1).

Factor II

37. To like to make fun of people who do things you regard as stupid. (Agg.) .62
35. To like to use words which other people often do not know the meaning of. (Exh.) .58
28. To like to criticize someone publicly when they deserve it. (Agg.) .30
33. To like to have your friends make a fuss over you when you are hurt or sick. (Suc.) .29
17. To like to be the center of attention in a group. (Exh.) .25

This dimension might be labeled Aggressive Arrogance or Criticalness. The presence of Items 33 and 17, however, suggests that arrogant and critical behavior might possibly be considered desirable in this instance because of the social attention thereby derived

and the attendant gratification of certain narcissistic and exhibitionistic needs. An alternative and perhaps more descriptive factor label, then, might be Narcissistic Exhibitionism.

Factor III

16. To like to have friends who feel sorry for you when you are sick. (Suc.) .67
 33. To like to have your friends make a fuss over you when you are hurt or sick. (Suc.) .51
 17. To like to be the center of attention in a group. (Exh.) .48
 11. To like to have your friends confide in you and tell you their troubles. (Nur.) .39
 15. To like to participate in new fads and fashions. (Chg.) .35
 35. To like to use words which other people often do not know the meaning of. (Exh.) .31
- 8, 6, 39, 9, and 19 also have loadings above .25.

This dimension, which appears to be somewhat related conceptually to Factor II as a different aspect of ego-supporting needs, might be labeled Emotional Dependence or perhaps Succorance. In this connection Item 15 is probably not viewed so much as an indication of the need for change as of the need for attention and support, so that in this case, as on some of the other factors, one need may be instrumental to the gratification of another. Thus the need for change may here be activated in the service of the need for succorance by *subsidiation* (Murray, 1938, p. 86), a process not considered in the development of scales for one need at a time.

Factor IV

20. To like to listen or to tell jokes in which sex plays a major part. (Het.) .64

24. To like to read books and plays in which sex plays a major part. (Het.) .59
 32. To like to be able to put in long hours of work without distractions. (End.) .25

This factor appears to be a doublet representing Sexual Interests. The appearance of Item 32, and perhaps even the next three items in order (28, 19, 38), is interesting from a psychoanalytic viewpoint. The absence of Item 13 dealing with "love for the opposite sex" suggests a possible differentiation in the desirability of love relationships and substitute sex interests.

Factor V

30. To like to attack points of view that are contrary to yours. (Agg.) .56
 40. To like to argue for your point of view when it is attacked by others. (Dom.) .54
 31. To like to do things your own way without regard to what others may think. (Aut.) .44
 42. To like to judge people by why they do something—not by what they actually do. (Int.) .36
 9. To like to tell other people about adventures and strange things that have happened to you. (Exh.) .25

This factor might represent something like Intellectual Aggressiveness or Active Intellectual Independence.

Factor VI

25. To like to write letters to your friends. (Aff.) .67
 39. To like to show a great deal of affection toward your friends. (Nur.) .41
 23. To like to have strong attachments with your friends. (Aff.) .31
 21. To like to stick at a job even when it may seem as if you are not getting anywhere. (End.) .27

1 and 5 also have loadings above .25.

This dimension, which might be labeled Interpersonal Involvement, reflects a viewpoint in which affiliative and affectionate behavior is judged desirable.

Factor VII

41. To like to move about the country and live in different places. (Chg.) .69
 6. To like to travel and see the country. (Chg.) .55
 29. To like to analyze your own motives and feelings. (Int.) .25

This factor appears to be a highly specific doublet reflecting Interest in Travel. The need for change may be involved in this dimension, but it is of interest to note that Item 15, dealing with participation "in new fads and fashions," received a low loading.

Factor VIII

26. To like to have your meals organized and a definite time set aside for eating. (Ord.) .52
 22. To like to be in groups where someone else takes the lead in deciding what you are going to do. (Def.) .43
 38. To like to have your life so arranged that things run smoothly and without any change in plans. (Ord.) .35
 1. To like to conform to custom and avoid doing things that people you respect might consider unconventional. (Def.) .35

This dimension might be labeled Compulsive Conformity. It reflects a viewpoint in which conforming deference, a kind of detached disinterest and orderliness are considered desirable.

Factor IX

32. To like to be able to put in long hours of work without distractions. (End.) .61
 27. To like to do your very best in whatever task you undertake. (Ach.) .56
 6. To like to travel and see the country. (Chg.) .51
 34. To like to be loyal to your friends. (Aff.) .48
 12. To like to be able to come and go when you want to. (Aut.) .46
 36. To like to treat other people with kindness and sympathy. (Nur.) .45
 29. To like to analyze your own motives and feelings. (Int.) .44
 4. To like to learn what great men have thought about various problems in which you are interested. (Def.) .40
 2. To like to start in and keep working on some assignment until it is completed. (End.) .39
 14. To like to be recognized as an authority in some job, profession, or field of specialization. (Ach.) .36

5, 1, 10, 40, 23, and 8 also have loadings above .25.

This large factor appears to reflect an Achievement-Oriented, Middle-Class Stereotype of desirable behavior or a kind of Protestant Ethic. This popular conception favors a hard-working (Items 32, 2), success-oriented (27, 14), self-broadening (6, 4), friendly (34, 36, 23, 10), personally independent (12 and perhaps 40), socially con-

ventional (1), orderly (5), responsible (8) orientation.

The Factors in Relation to the Group Desirability Scale

Median desirability ratings computed for the total sample are presented in Table 1 for each item. For correlational purposes, these median ratings are very similar to the more refined successive intervals scale values used by Klett (1957b). This median desirability scale was correlated with each of the nine factors in turn, and the coefficients were found to be $-.04$, $-.65$, $-.30$, $-.33$, $-.11$, $.12$, $.19$, $.15$, and $.76$, respectively. The multiple correlation for the nine factors in predicting the median values was $.96$; the β -weights were $-.05$, $-.44$, $.15$, $-.03$, $.24$, $.06$, $.29$, and $.78$, respectively. Thus, the vector of medians can be oriented in the nine-dimensional factor space in such a way as to account for almost all of its variance. This is not a necessary finding, since the 42 median values represent averages of 42 frequency distributions, and the means of these distributions were essentially eliminated in the present analysis of intercorrelations. However, one might expect the average desirability scale for the judging group to be related to the component dimensions, particularly to the larger or dominant factor, since a point of view prevalent in the sample under study should make a larger contribution to the average scale.

Response Biases and Possible Artifacts

In the analysis of ratings based on a fixed number of categories, there are several possible sources of bias which could produce artifacts in the factor space. For example, ratings of items with medians near either extreme of the scale must be skewed, and this differential skewness has a small but consistent effect on the corresponding correlation coefficients, which could generate spurious dimensions similar to the "difficulty" factors in mental test analysis. Such factors should be consistently related to extreme median values, as are Factors II and IX in the present analysis; the other seven factors, however, involve mostly items from the same upper-middle range (Categories 5 to 7). Factor II includes some, but not all, items near the undesirable

extreme and correlates $-.65$ with the medians; Factor IX involves mostly desirable items and correlates $.76$ with the medians. However, if a rotation is made in the plane of Factors II and IX, essentially equating IX with the medians, the new Factor II would then include primarily Items 35, 37, and 32, which are fairly consistent psychologically with the old Factor II but which involve both extremes of the rating scale. This leaves Factor IX highly related to the medians, and possible spurious influences should be considered in interpreting it. However, it is doubtful that differential skewness would generate the largest factor, especially when skewness was not extreme in the present data: the rating distributions for the 42 items selected had no modes in either extreme category and only 14 modes in either Category 2 or 8.

Another source of potential bias stems from consistent rater preferences for specific categories, such as a predominant use of the extremes or the middle range in making judgments. Most of the raters in the present sample, however, appeared to use a wide range of categories, with only about a dozen Ss showing noticeable preferences for specific categories or restricted ranges. Although no more than two of these Ss appeared to have markedly similar response patterns, consistent but slight category biases operating throughout the sample could distort the factor space and should be considered in the interpretation. In this connection, it is interesting to note that, when Factors II and IX are rotated as mentioned above, the new Factor II has high loadings for items from both extremes and could reflect intensity of rating.

Such rating biases thus offer alternative interpretations for Factors II and IX, but since there is little evidence for their occurrence in the present data, the simple structure orientation is preferred.

DISCUSSION

The finding of divergent viewpoints about social desirability is obviously highly dependent upon the sample of individuals studied. For example, one could probably assemble a group of people with sufficient homogeneity that their ratings of desirability would reflect only one consistent point of view. The issue

in question, however, is not whether a single scale can be obtained for some particular population, but whether or not such an average scale is of sufficient generality to use in personality measurement to control for the tendency to respond in a socially desirable way. The present finding of several different dimensions of desirability suggests that such an average scale would not provide an adequate control in individual assessment and would have inconsistent effectiveness from subject to subject. Equating in terms of a large dimension which seemed to reflect a popular set of values for Western culture, such as Factor IX in the present analysis, might prove adequate for studying interrelations among variables at the *group* level; but for *individual* assessment the controls could be quite ineffective, particularly for people whose perceptions of desirability differed markedly from the group consensus.

The multidimensionality of social desirability ratings, the major finding of the present study, naturally must be cross-validated on normal populations. The stability and generality of the particular nine factors found here would be of interest, but it might be expected that several new and different points of view would be uncovered by varying the type of subjects sampled and items rated. However, it must also be considered that wide group differences could occur with only minor attendant changes in a factor structure which spanned several points of view. Morris and Jones (1955), for example, found that, although the ratings of various value orientations differed considerably across cultures, the factors obtained proved to be very similar.

In a society with widely differing social and subcultural influences, it is not surprising that there should be several dimensions of social desirability. However, the present finding of as many as nine factors, on a fairly restricted sample of 42 items and 108 cases, suggests the possibility that the consistent variation is not so much among clearly defined social groups, such as high school students or basic airmen, as among individual points of view. It is true that this diversity in viewpoint might be related to peculiarities in the judging sample, but in that case the factors could be important in their own right in relation to psychopathology. However, as

Edwards (1957b, p. 48) has noted, his instructions for rating social desirability are similar to Rosen's (1956) for personal desirability, in that the *S's own* judgments of the desirability of traits in others are emphasized. It might be expected that individual differences in such personal views of desirability would generate many factors, even for relatively homogeneous, normal populations. Thus, since the term "social desirability" appears somewhat misleading, even when discussing Edwards' (1957b) results, it would seem appropriate to substitute another label (cf. De Soto, Kueth, & Bosley, 1959; Rosen, 1956) which emphasizes differences in individual viewpoints of desirability.

The discussion thus far has emphasized the reduction of error variance in personality measurement by controlling the effects of the tendency to respond in a socially desirable way. This response tendency, however, may also be viewed as a stylistic personality variable (Jackson & Messick, 1958), and in this connection, rather than trying to eliminate the effect, attempts might instead be made to measure the individual's position on the different dimensions. If, for example, measures of differential tendencies to respond in various desirable ways can be formulated, they may be used on the one hand as personality variables in their own right and on the other as multiple control measures to partial differential desirability biases out of content scales. Since a reliable tendency to respond in desirable ways might in itself reflect certain traits or needs in the respondent, much research is needed to develop control techniques (cf. Helmstadter, 1957), which do not require, as partial correlation does, that the corrected content scale be uncorrelated with the response tendency. It must also be remembered that the factors discussed in the present paper represent dimensions of *judged* desirability, which may or may not be related to dimensions of the tendency to *respond* in desirable ways, if such exist.

The proposed treatment of desirability dimensions as personality variables is similar in many ways to the diagnostic use of "ideal-self" descriptions, which have been widely employed in evaluating outcomes of psychotherapy (Rogers & Dymond, 1955). Osgood, Suci, and Tannenbaum (1957, Ch. 6) have

also utilized ratings on the Semantic Differential, which has a large evaluative component, to compare conceptual structures in different personalities and to study personality change during psychotherapy. Incidentally, this study of differences in individual conceptions appears fruitful even though comparisons at the group level show nearly identical factor structures for different cultures and pathologies (Osgood, Suci, & Tannenbaum, 1957, pp. 170-175). Thus, the presence of consistent individual differences in the perception of desirability offers the possibility that the seemingly objective task of rating the desirability of a trait may be useful in making inferences about the personality of the rater.

SUMMARY

Attempted controls in personality inventories for the tendency to respond in a socially desirable way usually assume that the domain of desirability is unidimensional. The present study attempted to evaluate this assumption and to explore the nature of any multidimensional structure uncovered. Ratings of the desirability of personality items from the Edwards Personal Preference Schedule were intercorrelated and factor analyzed. Nine factors were extracted and rotated to oblique simple structure. The factors were labeled and interpreted in terms of different "points of view" as to what is desirable. The finding of several dimensions of desirability suggests that the equating of items in terms of a single average desirability scale would leave much of the consistent variance of the ratings uncontrolled. Such a control method would provide inconsistent effectiveness for individual assessment, depending upon the S's conformity to the group consensus of desirability.

REFERENCES

- DE SOTO, C. B., KUETHE, J. L., & BOSLEY, J. J. A redefinition of social desirability. *J. abnorm. soc. Psychol.*, 1959, **58**, 273-275.
- EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90-93.
- EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corp., 1954.
- EDWARDS, A. L. *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts, 1957. (a)
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957. (b)
- FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171-175.
- FRENCH, ELIZABETH G. A note on the Edwards Personal Preference Schedule for use with basic airmen. *Educ. psychol. Measmt.*, 1958, **18**, 109-115.
- HANLEY, C. Deriving a measure of test-taking defensiveness. *J. consult. Psychol.*, 1957, **21**, 391-397.
- HEINEMAN, C. E. A forced-choice form of the Taylor Anxiety Scale. *J. consult. Psychol.*, 1953, **17**, 447-454.
- HELMSTADTER, G. C. Procedures for obtaining separate set and content components of a test score. *Psychometrika*, 1957, **22**, 381-393.
- JACKSON, D. N., & MESSICK, S. J. Content and style in personality assessment. *Psychol. Bull.*, 1958, **55**, 243-252.
- KLETT, C. J. The stability of the social desirability scale values in the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1957, **21**, 183-185. (a)
- KLETT, C. J. The social desirability stereotype in a hospital population. *J. consult. Psychol.*, 1957, **21**, 419-421. (b)
- KLETT, C. J. The social desirability stereotype and some measures of psychopathology. *J. consult. Psychol.*, 1957, **21**, 450. (c)
- KLETT, C. J., & YAUKEY, D. W. A cross-cultural comparison of judgments of social desirability. *J. soc. Psychol.*, 1959, **49**, 19-26.
- MORRIS, C., & JONES, L. V. Value scales and dimensions. *J. abnorm. soc. Psychol.*, 1955, **51**, 523-535.
- MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana, Illinois: Univer. Illinois Press, 1957.
- PINZKA, C., & SAUNDERS, D. R. Analytic rotation to simple structure: II. Extension to an oblique solution. Res. Bull. 54-31. Princeton, N. J.: Educational Testing Service, 1954. (Multilith)
- ROGERS, C. R., & DYMOND, R. F. (Eds.) *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1955.
- ROSEN, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.*, 1956, **52**, 151-158.
- SISSON, E. D. Forced-choice—the new Army rating. *Personnel Psychol.*, 1948, **1**, 365-381.
- THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. Chicago Press, 1947.
- TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.
- WAHLER, H. J. Social desirability and self-ratings of intakes, patients in treatment, and controls. *J. consult. Psychol.*, 1958, **22**, 357-363.

SOME CONCEPTS CONCERNING THE MENTAL HEALTH OF THE INDIVIDUAL¹

DONALD C. KLEIN

Human Relations Service of Wellesley, Massachusetts

The concept of mental health defies definition despite efforts to render it operational and objective. The intent of this paper is to suggest an alternative set of concepts, which may prove more useful to the growing field of research and technology committed to mental health pursuits. The need for such concepts is clear. The urge of man to mold and perfect life has in the past two decades shifted, particularly in Western lands, from an attack on sociopolitical and economic forces to the intrapsychic and interpersonal forces believed to prevent the attainment of "the good life" (Duhl, undated). An example of this is given in the following quotation from a brochure issued by a mental health association in one of the northeastern states: "IF EVERYONE ENJOYED MENTAL HEALTH, there would be no crime, no insanity, and no deepseated unhappiness born of a feeling of frustration or of a sense that one does not belong." It would appear that we have by now badly overloaded the concept of mental health to the point where it has become virtually useless either for action or research.

Much of the early epidemiological work in this field has been concerned with cases of demonstrable pathology, i.e., patients accepted for hospital or outpatient care. From the viewpoint of such studies, mental health is the absence of mental illness, but even here there is the problem of determining generally acceptable criteria for deciding what is or is not an instance of genuine pathology. Problems multiply when we turn our attention

¹ While maintaining responsibility for the assertions contained herein, the author wishes to express appreciation to Erich Lindemann, whose perspective and great vision have led to this attempt at clarification. Gratitude is also given to Frank Babbott for assistance with epidemiological concepts, to Frank and Joan Babbott for editorial help, and for their critical comments and suggestions to L. Borstelmann and W. Hollister.

from hospitals and clinics to the general population. The few statistics regarding distribution of problems in the general population seriously challenge our present inadequate concepts of mental health and emotional illness. For example, the "psychopathology of everyday life" has been demonstrated recently by Leighton (1956), who found that psychiatric appraisal of a sample of the population of a Canadian town revealed psychiatric impairment among at least 37% of the citizens.

Jahoda (1955) and Lindemann (1953), among others, have laid down explicit criteria for positive mental health. Their definitions describe the mental health of the individual in a social context. They seek to avoid enclosing him in the strait jacket of statistical normality or requiring him to acquiesce to malign environmental circumstances. In their definitions, adequate perception of reality, ability to adjust to inevitabilities and to master resolvable difficulties, and—for Lindemann—attainment of personal satisfactions and tranquility without impairment of others' functioning are viewed as prime characteristics. Thus, mental health is defined not as the absence of symptoms, but rather as the presence of three qualities in sufficient abundance: good reality testing, active mastery of the environment, and a feeling of contentment not based on victimization of others. At this point we are left with many questions regarding the interrelations of the three qualities. Does a deficiency in one imply that an individual no longer enjoys good mental health? Or must all three be reduced for ill health to ensue? Is any one—reality testing, for instance—a prerequisite for the other?

Such questions arise, in part, because three different basic meanings have been attached to the concept of mental health. In order not to perpetuate such confusion, the balance of this paper will concern itself with

the three connotations of mental health, using a different term for each. First, there is that meaning attached to the term "mental health" which appears to refer to the more enduring condition of the personality (hereafter referred to as "soundness"); second, there is that meaning which appears to refer to the more immediate state of health, equilibrium, or current effectiveness of the person in terms of the reaction of the organism to a stressor agent (hereafter referred to as "well-being"); third, there is that meaning which appears to refer to the ability of the individual to avoid illness even when exposed to illness producing agents (hereafter referred to as "emotional stability").

SOUNDNESS

By "soundness" is meant the level of integration of the general, more enduring personality structure. It is assessed in the perspective of the culturally shared ideal of the optimal in growth for any age level or other significant role dimension. Thus, in our culture it might include such attributes as adaptability, well developed social skills, a high degree of initiative, a sense of social responsibility, and a general perfusion of modest optimism in one's perception of man's struggle with nature and himself.² It is believed that such ideal attributes, though culture bound, tend to vary inversely with severity of psychiatric pathology. Perhaps psychiatric nosology is itself culture bound. Possibly such ideal characteristics are indeed hallmarks of emotional maturity, reflecting the superiority of our culture over others. It seems more probable, however, that in this society—as in most, if not all, others—ideal types have emerged which exemplify one or more of the *several sound integrations* of which man is biologically capable as he attempts to satisfy needs, pattern behavior, form social groupings, and find secondary satisfactions in interpersonal contacts.

Certain possible universal *attributes* of integration suggest themselves. The writer believes the attributes of an individual with

a high soundness level can be expressed in his own terms pretty much as follows: "I feel I belong; I am a worthwhile person; I know where I stand." The external observer might choose to express the first attribute in terms of the clarity of the social role of the individual and the value ascribed to that role by the individual and others. He might then speak of the personal sense of worth of the individual and add an assessment of the appropriateness of role behavior as compared with the expectancies of the significant reference group.³

WELL-BEING

The concept "well-being" refers to the current state of equilibrium between the individual and the social-emotional environment impinging upon him. It is postulated that almost all individuals exhibit virtually continuous fluctuations in well-being. It is further believed that nearly all individuals undergo repeated states of ill-being and that such states are manifested in one or more of the commonly recognized patterns described in articles on combat fatigue, disaster fatigue, or psychological shock reactions observed in air raids and other mass stress situations (e.g., Janis, 1951). Finally, it has been noted in work with individuals undergoing emotional crises that states of ill-being tend to be time limited, the tendency being to return to a state of equilibrium whether or not treatment is available. The time period appears to depend upon the nature, severity, and duration of the stress.

Those who cannot arrive at an equilibrium following a state of ill-being are apt to suffer a serious decrement in soundness. This observation is supported by experience with

³ Some reinforcement for such speculations is offered by Schutz (1958), who suggests, on the basis of a large number of studies of group behaviors, that three core tasks in almost any interpersonal context relate to the questions: (a) Do I have or can I gain membership? (inclusion); (b) What is the structure or pattern of relationship into which I must fit? (control); and (c) To what extent can I safely express my personal needs in this relationship? (affection). Schutz further presents a schema in which each of the three interpersonal need areas is seen to vary according to whether it is handled in an "ideal" fashion, is "deficient" or "excessive" to a moderate degree, or is dealt with pathologically. Thus a pathological underdevelopment is believed to accompany schizophrenia.

² A somewhat similar conception is presented by Shoben (1957) in a model of integrative adjustment. He also includes self-control which, as will be seen, this writer believes is more appropriately considered separately as one characteristic of stability.

combat fatigue cases during World War II, who developed a higher proportion of chronic emotional problems when removed permanently from combat units as compared with similar cases which were quickly returned to their units after a brief respite during the Korean conflict. Therefore a mental health program concerned with prevention and health maintenance is justified in devoting itself to the brief care of those suffering temporary impairment of well-being in an effort to forestall reductions in soundness levels.

When is a condition of ill-being in a *population* serious enough to warrant attention? This is one of the most challenging questions facing the field of community mental health. It can be satisfactorily answered only when base or "endemic" rates of prevalence and incidence for specific states of ill-being have been established among major population groupings. Meanwhile, many of those working with more or less closed communities, such as mental hospitals, schools, and prisons, are usually very clear in their own minds when institutional tensions affecting the well-being of large numbers in their care have reached "epidemic" proportions.

In order to establish base rates, it is necessary for there to be something to count. In one study, which led ultimately to a clearly defined preventive program, it was possible to count the number of withdrawals of students from nursing training. Generally acceptable indices of emotional distress are not always so readily available. And yet common sense and current societal preoccupations suggest that they do exist. Thus, for example, acts of violence, such as murder or assault, should justifiably be counted as instances of poor mental "health" (i.e., emotional ill-being) whether or not they occur in individuals of high or low soundness. Rates of occurrence of such behavior may also be taken as indications of the general mental "health" (well-being) of a population. Such indices may be of greater import in the evaluation of mental health programs than figures concerning cases treated in and out of the mental hospital. So it is important to study acute states of ill-being or crises *on their own merits* (irrespective of degree of more central personality dysfunction) in order to learn something about their etiology

in the individual and their frequency and effects in the population.

EMOTIONAL STABILITY

The distinction between general soundness and well-being, the latter viewed as response to environmental stress, is derived from the ecologic notion that health or illness are states of equilibria between an organism and its environment (Burnet, 1953). The ability of the individual to cope with stress determines whether or not well-being is affected. While emotional well-being can be maintained under either ordinary or especially stressful circumstances, ill-being can result only from untoward reaction to stress. *Stability thus becomes the ability of the individual to cope with specific environmental stresses while maintaining a state of emotional well-being.* Stability levels of individuals can be compared inferentially by noting their behaviors under similar stressor conditions. Ultimately it may be possible to estimate stability levels by psychological laboratory tests, which have been found to correlate highly with adequacy of response to the specific stress or class of stresses in question.

As Whitehorn (1956) emphasizes, individuals differ markedly from one another in respect to situations they *perceive* as stressful. Given such variability, a phenomenological approach to the concept of stress suggests itself. Nevertheless, it is believed preferable for present purposes to consider stress in terms of forces impinging upon the individual, without regard to his reaction. Such a differentiation of stress from the organism's reaction permits clear-cut identification and ultimate categorization of psychological stressor agents. In preventive work the presence or absence of the stress in a population is worthy of study, apart from the reaction in any particular case.

Borrowing from the field of epidemiology, we have attempted to distinguish between four types of emotional stability.⁴

General Stability

The "general stability level" refers to the capacity of the individual to maintain equi-

⁴ In epidemiology it is usual to think of immunity in four categories: inherited, general immunity; acquired, general immunity; inherited, specific immunity; and acquired, specific immunity.

librium or well-being under conditions of stress without regard to the nature of the stress itself. A "specific stability area," on the other hand, refers to the ability to withstand a specific kind of emotional stress. Any specific area of stability, therefore, must be defined in terms of some specific stressor agent. Specific stability is discussed more fully below.

Much work needs to be done in order to arrive at an adequate appraisal of the ingredients of the general stability level. In this regard, it should be worthwhile to examine the possible general "immunizing" effects of certain kinds of prolonged, insight-oriented psychotherapies. It may ultimately be demonstrated, for example, that classic psychoanalysis represents not only a therapy, aimed at relieving a specific state of emotional ill-being, or psychological exercise, which raises the general soundness level, but also a form of "immunization," which helps the individual ward off a wide variety of potentially noxious stimuli by increasing general stability or supporting specific stability areas.

Evidence has been advanced (Eysenck, 1952), which suggests that recovery or improvement rates are roughly 60% regardless either of the nature, depth, or length of the treatment or of whether it is carried out by psychoanalysts or family physicians (who may prescribe a few weeks in the general hospital for rest and recuperation). Such comparisons obviously suffer from the fact that uniform criteria of successful treatment do not exist. Nevertheless, once such criteria are available, it is not outside the realm of probability that so-called insight therapies will be found to resemble many other palliatives in the degree to which they restore emotional well-being, but will be proved significantly better able to render patients more generally stable under subsequent stress conditions, with the result that such patients are less apt to suffer recurring states of emotional ill-being.

The distinction between general stability and soundness is fuzzy at best. It is believed that soundness has a moderately high positive correlation with general stability but that the relationship is not univocal. It is assumed that high general stability can exist in an individual of only moderate soundness, especially when the personality defenses of the

individual are such as to insulate him from the full impact of stress. Such defenses may not always be considered to reflect a high soundness level though they may be very effective aids to general stability. They include such reactions as repression, denial, selective inattention, and possibly even an ability under some circumstances to resort to the use of phantasy when reality becomes too bleak. Examples of apparently effective use of both denial and phantasy have been observed in some adult polio patients in respirators. These individuals appear to avoid the full impact of their predicament, whereas other patients not capable of so doing display depression and other signs of being overwhelmed by the all-too-real catastrophe.

Specific Stability

Studies of populations undergoing marked shifts in sociocultural environments may shed increasing light on the relationship between general and specific stability levels. It seems probable that individuals with high general stability will demonstrate areas of low specific stability when exposed to certain forms of stress. Careful review of soldier and civilian populations in wartime, of forced migrations of whole peoples, and of groups known to be either upwardly or downwardly mobile from a socioeconomic standpoint may suggest fruitful lines of future inquiry.

In order to study areas of specific stability, it is necessary ultimately to arrive at some categorization of types of stress conditions to which large segments of the population are routinely exposed. Some suggestions are to be found in the literature. Birth trauma is given some emphasis by certain psychoanalytic theorists. Perhaps here is the biologic prototype of the entry of the individual into a new, unfamiliar, confusing, and potentially harmful setting. Reactions to frustration or interference with goal-directed behavior have also been studied in the laboratory and find their counterparts in clinical patients, who become anxious or depressed when advancement up the economic or social ladder seems no longer possible. Lindemann's (1944) observations of reactions to bereavement have led to a formulation of emotional hazard which emphasizes sudden alterations or disruptions in key role relationships.

Inherited Stability

The distinction between inherited and acquired immunity—however useful for public health—may seem simply another way to raise the hackneyed controversy of nature vs. nurture, heredity vs. environment. It may be possible, nonetheless, to make some meaningful distinctions without promoting further useless controversy. At the outset, it may be helpful to remind ourselves that behavior does not exist in the absence either of the organism or its environment.

There are those (e.g., Erikson, 1950) who suggest that children are born with temperamental response tendencies (heredity, if you will) which affect parental responses to them (i.e., environment) in various ways, depending in part upon what the mother herself brings to the relationship. If the mother's initial attempts at mothering are discouraged by certain response tendencies in the child, permanent damage may be done to the affectional ties between them. A recent study by Klebanoff (1959) supports this viewpoint. He found that mothers of children with brain damage and of schizophrenic children responded similarly to questions on a child rearing attitude scale and that both, in turn, showed significantly more pathological attitudes toward child rearing and the family than did mothers of normal children.

If some children are born with response tendencies rendering them highly vulnerable to maternal rejection, they may be described as having inherited a kind of emotional instability and therefore as being susceptible to subsequent emotional maladjustment. Impairment of the mother-child relation as a result of the way in which the mother handles the child's inadequacy will result in a child with an extremely low soundness level (Fries, 1944; Mittelman, 1954). Obviously such speculations regarding inherited stability and instability must be tested ultimately by the detailed observations of newborn children, such as those of Fries (1944) who correlated response tendencies of neonatal infants with the specific mothering patterns required by each type of child to insure healthy psychological development.

Acquired Stability

In epidemiology the concept of acquired immunity implies that the organism has had

some experience with the stressor agent and, in the process, has developed successful (i.e., nonillness producing) ways of coping with it for some period of future time. During the period of immunity, the organism and the agent are able to coexist without undue harm to the former, or the invader is destroyed.

In this area, as in so many instances where we are concerned with the relationship between sickness and health, it is desirable to begin first with the observations of groups of people who have become sick under demonstrable stress conditions. Having noted the reactions of those who become markedly ill, it should become possible to shift our attention to those whose emotional well-being is only slightly or temporarily impaired by similar stress. Finally, it may be possible to discover situations in which small amounts of stress render individuals increasingly able to cope with more massive doses of similar stresses in the future. The following statement by Frank (1957) from a national magazine expresses the belief that sibling rivalry, known in its virulent form to play a part in disturbing emotional equilibrium, may under certain conditions act so as to confer a kind of stability in a specific area:

In families where each child feels he has reasonable access to his parents' attention and affection, sibling rivalry is not only inevitable but desirable. It is inevitable because every child, at least initially, wants to be his parents' favorite. It is desirable because every child is bound to meet conflict sooner or later, at school if not before, and there is no better place in which to learn how to cope with it constructively than in the home where there is love and concern for both parties. Sibling rivalry is one of the major ways by which children can develop a sense that the world is, within limits, a trustworthy place in which they can feel reasonably secure (p. 59).

This point of view must be examined and carefully tested within the framework of specific responses to specific stress conditions. Otherwise, there is the possibility that it will gain popular support prematurely and join the growing list of changing fashions in child rearing. The danger of such uncritical acceptance is readily seen after consideration of the probability that, in some circumstances, small stresses do not immunize, but rather cumulate and eventually overwhelm the individual.

In the search for specific areas of acquired

stability, initial priority must go to the search for stressor agents, a search that is well underway. Ultimately, however, it becomes necessary to relate major stress areas, which have been shown to produce casualties, to prototypical situations in the early life development of the organism that occasion mild stress without disabling the child. For example, while all of us learn in a variety of ways a multiplicity of social skills, relatively little has appeared in the literature regarding such skills and how they are established.⁵ In this search, it may be desirable to examine a wide variety of disturbance of parent-child relationships noted in the child guidance literature because each may have its counterparts in the repertoire of role interactions of most, if not all, parents and children. Such a search might have two desirable results: first, it might suggest conditions under which stressor agents exceed a safe tolerance level; second, it might point to conditions under which individuals are not exposed to sufficient amounts of stress, thus rendering them less stable in similar situations in later life when more enduring personality patterns have been developed.

SUMMARY

This discussion has put forth the suggestion that the search for definitions of mental health be abandoned, at least temporarily, in favor of a differentiated approach to the problem of human adjustments and maladjustments. It has been suggested that current concepts of mental health and illness are primarily concerned with the long standing emotional or psychological condition of the individual, for which the term "soundness" is proposed. The immediate state of "well" or "ill-being" of the individual at any particular time is proposed as suitable for more intensive study, both because of its possible effect on the soundness level in some instances, and because it should be considered in and of itself as a period of acute illness or malfunctioning and worthy of note in population studies. The vulnerability of the individual to environmental stress has been

discussed in terms of a possible "general stability factor," related to but not identical with soundness level, and of areas of "specific stability" which offer a promising field for further study in terms of possible implications for child rearing and educational practices. Perhaps by following this more differentiated approach to mental health, we may find that apparently conflicting doctrines of today may be viewed as much less contradictory in a more adequate viewpoint of tomorrow.

REFERENCES

- BURNET, F. M. *The natural history of infectious disease*. (2nd ed.) Cambridge: Cambridge Univer. Press, 1953.
- CAMERON, N. *The psychology of behavior disorders*. Boston: Houghton Mifflin, 1947.
- DUHL, L. Psychiatry and the community. Unpublished Manuscript, National Institute of Mental Health, undated. (Mimeo.)
- ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.
- EYSENCK, H. J. The effects of psychotherapy: An evaluation. *J. consult. Psychol.*, 1952, 16, 319-324.
- FRANK, J. D. Are you a guilty parent? *Harper's*, 1957 (April), 214, 56-59.
- FRIES, M. Psychosomatic relationships between mother and infant. *Psychosom. Med.*, 1944, 6, 159-162.
- JAHODA, MARIE. Toward a social psychology of mental health. In A. M. Rose (Ed.), *Mental health and mental disorder*. New York: Norton, 1955.
- JANIS, I. *Air war and emotional stress*. New York: McGraw-Hill, 1951.
- KLEBANOFF, L. B. Parental attitudes of mothers of schizophrenic, brain injured and retarded, and normal children. *Amer. J. Orthopsychiat.*, 1959, 29.
- LEIGHTON, D. C. The distribution of psychiatric symptoms in a small town. *Amer. J. Psychiat.*, 1956, 112, 716-723.
- LINDEMANN, E. Symptomatology and management of acute grief. *Amer. J. Psychiat.*, 1944, 101, 141-148.
- LINDEMANN, E. Mental health—fundamental to a dynamic epidemiology of health. In I. Galdston (Ed.), *The epidemiology of health*. New York Acad. Med., 1953.
- MITTELMAN, B. Motility in infants, children, and adults. *The psychoanalytic study of the child*, 1954, 9, 142-177.
- SCHUTZ, W. *FIRO: A three-dimensional theory of interpersonal behavior*. New York: Rinehart, 1958.
- SHOBEN, E. Toward a concept of normal personality. *Amer. Psychologist*, 1957, 12, 183-189.
- WHITEHORN, J. C. Stress and emotional health. *Amer. J. Psychiat.*, 1956, 112, 773-781.

(Received July 7, 1959)

⁵ One notable exception to this statement is Cameron's (1947) work on the social origins of mental illnesses, which suggests that emotional breakdown is essentially founded upon inadequate or inappropriate social perceptions and skills.

EGO STRENGTH AND CONFLICT DISCRIMINATION: AN EXPERIMENTAL CONSTRUCT VALIDATION OF THE EGO STRENGTH SCALE¹

MAURICE KORMAN

University of Texas Southwestern Medical School

Barron's *Es* scale was validated against the predictive criterion of successful outcome in psychotherapy (Barron, 1953). Because of the method of its construction, its content and correlates, it is suggested that it measures ego strength, a presumably latent trait often obscured by the vividness of psychopathology. Research with the *Es* scale has tested its criterion validity by examining its predictive and postdictive power (Wirt, 1955, 1956) and investigated its construct validity principally by computing its correlation with indices based on other psychometric instruments such as the Rorschach and the Bender-Gestalt (Tamkin, 1957; Tamkin & Klett, 1957). This type of "psychometric" construct validation presents a number of difficulties. First, such tests frequently can marshal but little empirical or theoretical evidence for being acceptable operational specifications of the construct in question, a drawback not remedied by a plethora of low but positive correlations (see Jenkins & Lykken, 1957). Secondly, as Campbell and Tyler (1957) have pointed out, intertest correlations may represent common but construct-irrelevant features (e.g., response sets). They add: "Construct validation is epitomized by the correlation between two or more specifications of a construct maximally *different* in apparatus or method" (p. 91). In Cronbach's terms (1957), the present study attempts to incorporate two basically divergent modes of approach, the correlational and the experimental.

¹ This research was carried out at the Psychopathic Hospital of the State University of Iowa. Carl Bickert's help in running subjects is gratefully acknowledged.

A version of this paper was given at the Annual Meeting of the APA, Cincinnati, 1959.

Because the *Es* scale is empirically defined it is especially desirable for other specifications of ego strength to be linked to a promising theoretical network. A congruence between such measures would then go beyond indicating that certain face valid experimental procedures relate to the construct interpretation of the test; it would serve as a bridge relating the psychometric device to an existing theoretical network. Dollard and Miller's explication of ego strength in terms of higher mental processes (1950) was taken as the starting point. The components of higher mental processes are many; principally, they include the ability to appropriately mediate discrimination, generalization, transfer, learned drives, and reinforcement. In this study the capacity of an individual to resolve a discrimination conflict was chosen as the experimental index of ego strength. The reasons for this choice were twofold: first, there is an extensive literature beginning with Pavlov (1927) on the conflictful characteristics of difficult discriminations; secondly, since the increased ability to solve conflicts and to make appropriate discriminations are both among the commonly accepted goals of psychotherapy, a more felicitous parallel between the experimental procedures and the *Es* scale is achieved. In addition, the parameters of discrimination problems are sensitive to experimental manipulation; one can readily increase the conflict loading of a discrimination task by placing the subject (*S*) under instruction to be as fast as he can while at the same time informing him of the irrevocable nature of his decision once it is made.

The hypothesis was therefore formulated that a high *Es* group should be able to resolve

discrimination conflict more speedily than a low *Es* group. The differences between the *Es* groups should be greatest for the condition of extreme discrimination difficulty (the most conflict provoking) and decrease as the discrimination becomes easier. In the case of easiest discrimination some general disruption of performance might still be expected because of the generalization on conflict (Brown, 1942), although the prediction here is less clear.

PROCEDURES

A discrimination situation involving the method of constant stimuli was used, patterned after Festinger (1943) and Block and Petersen (1955). The *Ss* sat in a darkened room approximately 4½ feet away from a black box, behind which were two lights placed about 10 inches apart and shining through holes ½ of an inch. The top light (standard stimulus) was a 7½-watt opaque bulb, while the variable light was a regular 40-watt white bulb. The voltage applied through the variable light at 7 different rheostat positions was 60, 70, 78, 80, 82, 90, and 100 volts, respectively. In addition the voltages were chosen in such a way as to make the two lights nearly equal in intensity when 80 volts were applied at the variable light. The *S* sat behind a table with his preferred hand on a toggle. He was instructed to make a judgment as to whether the variable light was brighter or dimmer than the standard light and to indicate his choice by pushing the toggle in the appropriate direction. In order to increase the difficulty of the task, a .5-sec. interval occurred between the standard and the variable light, at the end of which a clock was activated which ran until the *S* pushed the toggle to the left or to the right. He was further told: "I'd like you to make up your mind as soon as it's possible for you but keep in mind that once you've pushed the lever up or down you can't change your mind again."

Each of the seven intensities of the variable stimulus were presented a total of 15 times by using three successive random orderings of the stimuli and by presenting this complete set of 21 trials five times.

Subjects

Forty-seven recently admitted psychiatric inpatients were used as *Ss*. Excluded were patients with a diagnosis or suspicion of organic brain disease, patients undergoing some type of radical therapeutic interference such as electroshock or insulin, and patients too uncooperative, agitated, or withdrawn to participate. The *Ss* (31 women and 16 men) were administered the MMPI and their *Es* scores obtained through the Hathaway and Briggs norms (1957). The entire range of *Es* scores rather than the extremes was used in this study so as to make it a more severe test of the scale's power. Because of the sex-linked speed elements of the

task, the male and female *Ss* were considered separately. A cutting score (*T* score) of 41, the median, was used to separate high and low female *Es* groups. Since the *Es* scores for the smaller sample of males appeared to be markedly skewed, the same cutting score was adopted for them.²

RESULTS

Prior to any analysis of the data, rules were laid down for the identification of the "easy," "intermediate," and "difficult" conflict situations based on *Ss*' individual performances. The criterion used involved each *S*'s consistency in labeling any one of the seven intensities of the variable stimulus brighter (or dimmer) than the standard stimulus. Those intensities of the variable stimulus which were most consistently seen as brighter (or dimmer) and which were furthest away from the point of objective equality (the middle intensity) were chosen as representing the least difficult discrimination for that *S*. Likewise, the most difficult discriminations were those which were most inconsistently seen as brighter (or dimmer) and which were nearest to the point of objective equality. The third category, that of intermediate difficulty, was defined by choosing those intensity positions which fell between the easy and difficult discriminations with reference to the criteria outlined above. Each of these three difficulty levels was represented by two intensities, 15 judgments being given to each. The mean of the 30 prejudgment latencies at each difficulty level was the basic data of this study. That no distortion of the data occurred through these procedures is shown by the fact that extreme positions were chosen as easy discrimination cells while the center positions tended to become difficult discrimination cells; the high and low *Es* groups did not differ in this respect.

As can be seen from Table 1 the high *Es* group holds a few more psychotics and is slightly older than the low *Es* group; these differences are far from significant (for diagnosis, chi square = .18 with 2 *df*; for age, *t* = .83 with 43 *df*).

² Due to circumstances beyond control, fewer subjects than desired were run. The statistical analyses were performed on females and males separately, the latter group serving essentially as a small replication sample.

TABLE 1
CHARACTERISTICS OF HIGH AND
LOW *Es* GROUPS

Variable	High <i>Es</i>	Low <i>Es</i>
Diagnosis ^a		
Psychotics	40%	33%
Neurotics	28%	33%
Personality Disorders	32%	33%
Age		
Mean	31.85	28.32
SD	13.90	13.39
<i>Es</i>		
Mean	52.25	31.10
SD	7.75	5.76

^a Based on $N = 43$; diagnosis on four patients was deferred.

The mean latency scores of the high and low *Es* groups on easy, intermediate, and difficult discriminations are shown in Fig. 1. The low *Es* group has a longer latency throughout, although the difference becomes more marked at the intermediate and difficult levels. Figure 1 combines male and female data because of the remarkable similarity of their separate curves within each *Es* group.

In order to test the significance of the differences between the curves of Fig. 1, exclusive of the easy discrimination category, the log transformed latency scores were subjected to separate analyses of variance for males and females after it had been concluded, by means of the Bartlett test, that

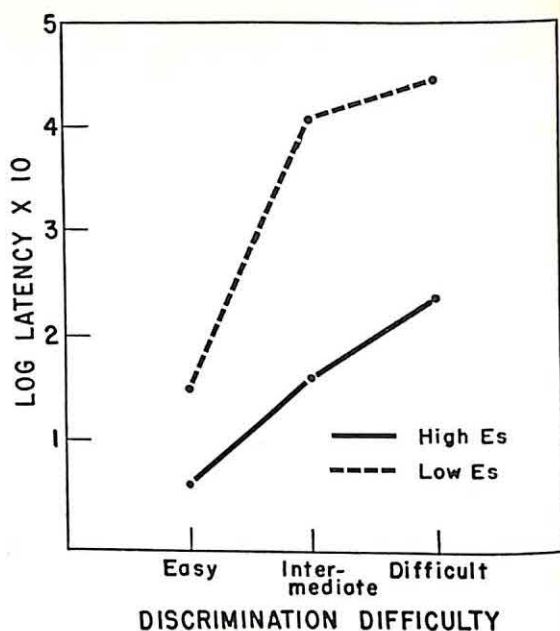


FIG. 1. Mean prejudice latencies of *Es* groups for increasing levels of discrimination difficulty.

the assumption of homogeneity of variance was not being violated. In order to compensate for the correlation introduced by repeated measurements on the same *Ss*, the analysis suggested for such cases by Edwards (1951) was used. Table 2 reports these results: in the case of both males and females there is a significant main effect between *Es* groups ($p = <.03$ for the females; $p = <.06$ for the males). The low ego strength patient takes longer in resolving the discrimination conflict than does the high ego

TABLE 2
ANALYSES OF VARIANCE FOR REPEATED MEASURES OF LOG TRANSFORMED LATENCY SCORES
FOR HIGH AND LOW *Es* GROUPS

Source	Females			Males		
	<i>df</i>	Mean Square	<i>F</i>	<i>df</i>	Mean Square	<i>F</i>
Between groups	1	.5618	4.94**	1	.3298	4.34*
Between <i>Ss</i> same Group	29	.1138		14	.0760	
Between Trials	2	.0674	5.14**	2	.0279	
Trials \times Groups	2	.0075		2	.0022	
Pooled <i>Ss</i> \times Trials	58	.0131		28	.0281	

* $p = .06$.

** $p = .03$.

TABLE 3

RESULTS OF *T* TESTS BETWEEN HIGH AND LOW
Es GROUPS AT THREE LEVELS OF
DISCRIMINATION DIFFICULTY

Discrimination Difficulty	<i>t</i>	
	Females (29 <i>df</i>)	Males (14 <i>df</i>)
Easy	1.05	1.78*
Intermediate	2.13**	1.87*
Difficult	2.10**	2.07*

* $p < .10$.

** $p < .05$.

strength patient. The differences between the two levels of discrimination difficulty were significant at the .03 level for the females but did not differ significantly for the males.

As Fig. 1 suggests, there is a disproportionate increase in latency for the low *Es* group as the discrimination difficulty moves from easy to intermediate and difficult as compared to the nearly linear relationship exhibited by the high *Es* group; the differences between easy and intermediate discriminations are significantly different for the two groups ($p < .02$) as are the differences between easy and difficult discriminations ($p < .03$). The groups do not differ, however, in their increment from intermediate to difficult discriminations. These findings generally received support from a comparison of the *Es* groups' performance on the three levels of discrimination difficulty. Table 3 reports the obtained *t*'s for males and females separately. In the case of the females, while the high and low *Es* groups do not differ on the easy discrimination, they differ significantly on both the intermediate and difficult levels of discrimination difficulty ($p < .05$). This trend is obscured somewhat in the case of the males because of the smaller *N*; the male high and low *Es* groups differ at the .10 level only with reference to all three levels of discrimination difficulty.

DISCUSSION

This study corroborates the findings of Tamkin (1957) and Tamkin and Klett (1957) with regard to the *Es* scale's inability to distinguish degrees of psychopathology as discriminated by psychiatric diagnosis. Con-

trary, however, to Tamkin and Klett's view that this is a "serious challenge to its validity," the present writer would suggest that this lack of association with overt psychopathology is a predictable and necessary by-product of the *Es* scale's construction which ignored the initial psychiatric status of the patients in favor of their treatment outcome. The difficulty, of course, lies in the surplus meanings of the term "ego strength" which lead to paradoxical conclusions.

The hypotheses concerning the relationship of the *Es* scale to discrimination conflict were generally borne out. In two separate samples, patients with low *Es* scores (a measure of ego strength, psychometrically defined) took longer than high *Es* patients to resolve a discrimination conflict (a measure of ego strength, experimentally defined). The latency between stimulus and response had been chosen as the dependent variable in this study because of its sensitivity to conflict instructions. The question arises as to whether the low *Es* group also made more errors in the successive discrimination judgments than the high *Es* group. Comparing the two groups on their total number of errors for the six intensities of the variable stimulus (leaving out the objectively "equal" center set) yields a nonsignificant *t* of .14 with 45 *df*. Since there is no a priori reason for suspecting personality variables to be related to the judgments in a psychophysical exercise, the null hypothesis can reasonably be accepted.

The *Es* scale's empirical derivation in the context of psychotherapeutic outcome had previously suggested a tentative identification of its underlying construct with ego strength, independently of any consideration of ego strength as one of the higher mental processes in the Dollard and Miller sense. The results of this study suggest that the "theory" underlying the *Es* scale is one which appropriately includes the construct ego strength as it is defined in the Dollard and Miller framework. Reciprocally, it is also possible to look upon these data as lending validity to an explication of ego strength patterned after Dollard and Miller and couched in terms of appropriateness of discrimination and conflict resolution. This can be maintained because part of the variance in such behaviors can be ac-

counted for by the variance of *Es* scores which are in turn associated with personality characteristics that make for a better "therapy bet."

With reference to the postulated relationship between *Es* scores and levels of discrimination difficulty, the results are likewise in a predicted direction although not in all respects. As the discrimination becomes more difficult, the low *Es* S's performance becomes disproportionately worse, while the high *Es* S's latency increases in a more stable, linear fashion. However, as Fig. 1 shows, the low *Es* S's latency on the most difficult discriminations is not very much greater than his latency for discriminations of intermediate difficulty. A number of possibilities present themselves. Low *Es* Ss possibly make only the distinction between quite easy and all other discriminations; this would suggest another type of inappropriate generalization on their part. It is also not improbable that their latency on discriminations of intermediate difficulties is close to the asymptote they can reach under the particular instructions given, i.e., to be both speedy and correct, and that a further increase in the difficulty of the discrimination has little effect.

SUMMARY

As a test of the construct validity of Barron's *Es* scale, 47 psychiatric inpatients were placed in a discrimination situation based on the method of constant stimuli. The hypotheses were (a) that a high *Es* group should be able to resolve a discrimination conflict more speedily than a low *Es* group and (b) that group differences should increase as the difficulty of the discrimination becomes greater.

The hypotheses were generally supported. The construct which the *Es* scale measures

is related to ego strength as defined in the Dollard and Miller framework in terms of capacity for appropriate discrimination and conflict resolution.

REFERENCES

- BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327-333.
- BLOCK, J., & PETERSEN, P. Some personality correlates of confidence, caution and speed in a decision situation. *J. abnorm. soc. Psychol.*, 1955, 51, 34-41.
- BROWN, J. S. Factors determining conflict reactions in difficult discriminations. *J. exp. Psychol.*, 1942, 31, 272-292.
- CAMPBELL, D. T., & TYLER, B. B. The construct validity of work group morale measures. *J. appl. Psychol.*, 1957, 41, 91-92.
- CRONBACH, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, 12, 671-684.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- EDWARDS, A. E. *Experimental design in psychological research*. Rinehart: New York, 1951.
- FESTINGER, L. Studies in decision: I. Decision time, relative frequency of judgements and subjective confidence as related to physical stimulus differences. *J. exp. Psychol.*, 1943, 32, 291-306.
- HATHAWAY, S. R., & BRIGGS, P. F. Some normative data on new MMPI scales. *J. clin. Psychol.*, 1957, 13, 364-368.
- JENKINS, J. J., & LYKKEN, D. T. Individual differences. In P. R. Farnsworth (Ed.), *Annual Review of Psychology*. Vol. 8. Palo Alto: Annu. Reviews, 1957. Pp. 79-112.
- PAVLOV, I. P. *Conditioned reflexes*. London: Oxford Univer. Press, 1927.
- TAMKIN, A. S. An evaluation of the construct validity of Barron's ego-strength scale. *J. clin. Psychol.*, 1957, 13, 156-158.
- TAMKIN, A. S., & KLETT, C. J. Barron's ego strength scale: A replication of an evaluation of its construct validity. *J. consult. Psychol.*, 1957, 21, 412.
- WIRT, R. D. Further validation of the ego strength scale. *J. consult. Psychol.*, 1955, 19, 444.
- WIRT, R. D. Actuarial prediction. *J. consult. Psychol.*, 1956, 20, 123-124.

(Received July 16, 1959)

THE DURATION OF THE ARCHIMEDES SPIRAL AFTERIMAGE IN THE DIAGNOSIS OF BRAIN DAMAGE¹

JEROME D. SCHEIN

Florida State University

Since Price and Deabler (1955) introduced the Archimedes spiral afterimage as a technique for differentiating brain damaged patients from other neuropsychiatric patients, there has been an increasing accumulation of evidence suggesting that the diagnostic efficiency of this test is far below that which clinicians would find useful as a routine screening device (Berger, Everson, Rutledge, & Koskoff, 1958; Gilberstadt, Schein, & Rosen, 1958; Goldberg & Smith, 1958; Philbrick, 1959). However, these investigators have relied on a pass-fail criterion for each trial, requiring only that the subject report whether or not he sees the negative afterimage (henceforth referred to as AIR for "afterimage report"). On the other hand, attempts to explain the failure of brain injured patients to report this phenomenon have implied that damage to the cortex reduces the *duration* of the afterimage (henceforth referred to as AID for "afterimage duration") with complete failure to report the afterimage representing an endpoint on the duration continuum.

Three theories have recently been proposed to account for the tendency of brain damaged individuals to report the negative afterimage induced by visual stimulation less frequently than patients with intact cerebrums. Saucer (1953) and Saucer and Deabler (1956) consider the perception of the afterimage to be

a function of a global effort of the entire cortex. Applying a principle of mass action, they conclude that brain injury would fore-shorten AID. Klein and Krech (1952) use the Pavlovian principle of inhibition to support their position that brain injury reduces AID, though their own research involved tactual aftereffects. Shapiro (1954), dealing directly with the perception of apparent movement, also applies the concept of inhibition in predicting that brain injury will fore-shorten AID.

Though the structure of their theories differ somewhat, these theorists are unanimous in predicting the effect of brain damage on AID. They provide the basis for avoiding the embarrassment of an all-or-none explanation of the earlier studies; i.e., one can account for brain damaged individuals who do give an AIR by assuming that they have a reduced AID, albeit not reduced below their perceptual limen. Thus, because duration is reduced, brain injured individuals should give fewer AIRs.

The empirical evidence on spiral AID is mixed. Eysenck, Holland, and Trouton (1957) asserted that some depressant drugs have effects isomorphic with brain injury, in that both increase inhibitory potentials in the brain. When they administered a placebo, a stimulant, and a depressant drug to 6 normal subjects (Ss), they found that the depressant drug (sodium amylobarbitone) significantly reduced AID, though the stimulant did not increase it. They felt that this result supported those theories of brain functioning which make use of the concept of cortical inhibition.

Three studies of brain injured patients

¹ This research is a portion of a PhD thesis submitted to the Department of Psychology, University of Minnesota. The author is indebted to William Schofield, Ephraim Rosen, Paul E. Meehl, and David T. Lykken of his dissertation committee for guidance and criticism. Appreciation is also due the staff of the Minneapolis VA Hospital for their unfailing cooperation and encouragement.

have included some observations on AID. Gallese (1956) noted that the false negatives (brain damaged patients who did report the afterimage) seemed to experience a less marked and apparently shorter AID than control patients. Page, Rakita, Kaplan, and Smith (1957) stated that they found no differences in AID between brain damaged and schizophrenic patients for two of the trials on which it was measured. In complete contradiction to these two studies, Spivack and Levine (1957) found that their brain damaged subjects reported *longer* AIDs than did a group of neurotic controls.

It is apparent that, while the theorists are in complete agreement, the small amount of evidence available is not. It was the purpose of this study to gather further data to help clarify the situation. Furthermore, if the theorists cited were correct, measuring the duration of the afterimage should increase the diagnostic efficiency of the test by identifying those patients whose brain injury does not reduce the afterimage to zero duration. These false negatives would be identifiable by their markedly reduced AID, even though they could report an afterimage.

METHOD

Subjects

An ongoing sample of all male veterans admitted to the Neurology and Psychiatry services of the Minneapolis Veterans Administration Hospital were tested within five days of their admission. This sampling procedure was adopted to control for chronicity of illness, eliminate patients undergoing shock therapy, and reduce the number of patients on a drug regimen. This method of selecting the sample also had the advantage of reducing errors which might have arisen had the experimenter had knowledge of the patient's diagnosis at the time of testing.

Of 129 cases admitted to the Neurology service 31 could not be tested: 3 expired, 17 were confined to their beds, 8 were too confused, 3 were discharged before testing. Seventeen additional cases were lost due to apparatus failure which occurred during the time the sample was being drawn. Of 50 admissions to the Psychiatry service, 5 could not be tested: 4 were confined to their beds, and 1 was discharged before testing.

Diagnosis

All diagnoses were final diagnoses arrived at by staff action. All cases having a diagnosis indicating central nervous system damage rostral to the for-

men magnum were further classified by a boarded neurologist² who determined whether or not the injury included damage to the cerebral cortex; if it did not, case was classified under subcortical brain damage. The psychiatric patients were divided into those having a psychotic diagnosis and those having a nonpsychotic diagnosis. Three cases who were tested left the hospital before diagnostic work-up could be completed. They are listed as "equivocal" in the section to follow. The remaining patients, being neither brain damaged nor psychiatrically ill, constituted the normal hospitalized control sample.

Drugs

Each patient's medication at time of testing was noted from the order sheet in his hospital chart. Two physicians and a pharmacologist³ then sorted the various drugs administered to patients in this study into four categories in accordance with their effect on the central nervous system within the range of therapeutic dosage. Drug Group 1 consisted of drugs having no effect on the central nervous system. Drug Group 2 contained all central nervous system depressants. Drug Group 3 included all central nervous system stimulants. Drug Group 4 consisted of any combination of drugs in Categories 2 or 3 which were being simultaneously administered to a patient.

Apparatus

An Archimedes spiral of 920° or 2½ turns about its center was constructed by painting a black spiral on a white disc 6 inches in diameter. The black and white lines were each ½ inch in width. The wheel was affixed to a reversible motor with a variable speed control which was set at 100 rpm.

Procedure

The S was seated 8 feet from the spiral. The illumination at the surface of the spiral was 16 foot candles, measured by a GE Exposuremeter, Model 8DW.

Two trials, one clockwise and one counterclockwise, were administered in accordance with Price and Deabler's (1955) instructions. Scoring was done in accordance with the rules established by Gilbertstadt, Schein, and Rosen (1958). If the patient failed on one or both of these trials to report the negative afterimage, he was given two additional trials. At the end of stimulation on these trials he was told to look at an orange wooden block placed next to the apparatus and to describe what he saw. If he then reported a negative afterimage correctly on two successive trials (i.e., the block appeared to move, grow, or change), he was administered the

² Grateful acknowledgement is made to Zondal Miller, consulting neurologist, for sorting these patients.

³ Grateful acknowledgement is made to Henry A. Johnson, Fred Miller, and A. Kleven for their assistance in classifying the drugs.

trials measuring duration. If he did not, he was given two further trials using a black circle drawn on white cardboard as the projection device. A patient who failed to report a negative afterimage on two successive trials of the six was considered to be diagnosed "brain damaged" by the test and no further testing was done.

As soon as two consecutive AIRs were given, the patient was instructed as follows:

What you just saw is called an afterimage; i.e., the line (block, circle) seemed to be moving even though you knew it wasn't. People with normal vision see these afterimages, so, in this sense your eyes are normal. Now I want to know how long

the afterimage lasts. To find this out, I will run the wheel several times just as before. But this time when it stops I want you to look at this block (circle). If you notice anything change, tell me when the change stops. If it doesn't change in any way, tell me that, too. I have this very accurate stop watch here, so I can tell just how long your afterimage lasts. Any questions? All right now, keep your eye on the wheel until I turn it off. Then immediately look at the block (circle). As soon as the block (circle) comes to a complete stop, you say "Now."

Eight trials were administered in the following order:

Direction	C	CC	C	CC	C	CC
Stimulation Time	15 sec.	30 sec.	30 sec.	15 sec.	15 sec.	30 sec.
Direction	CC	C				
Stimulation Time	15 sec.	30 sec.				

(C = clockwise; CC = counterclockwise)

The variations in time and direction were introduced to discourage a fixed set to say "Now" at the same time following each stimulation, regardless of what may have been perceived.

At the conclusion of the fourth trial patients whose first afterimages had been passed off on the block were instructed to look at the circle following stimulation and vice versa. Whether a patient was exposed to the block or the circle on his first four trials was determined by a letter which had been written alternately on each scoring sheet prior to testing.

Two test objects were used (a) to provide variety in the situation to help keep the patient's attention and (b) to test the hypothesis that brain damaged individuals would have longer AIDs for the three-dimensional block than for the two-dimensional circle. The latter notion was suggested by the fact that brain damage is thought to reduce the individual's ability to deal with the world in an "abstract" fashion, thus the three-dimensional block might be more logically capable of motion than the two-dimensional circle.

Duration was measured by allowing a split-hundredth watch to run continuously from the start of the trial until the patient indicated the afterimage had stopped. The AID was determined by subtracting the stimulation time from the total time elapsed. This method reduced watch-reading and "snap-back" errors.

The interval between trials was maintained at 15 sec. from the time the patient indicated the afterimage had ceased. On the AIR trials, intertrial time was 30 sec.

Patients who failed entirely to report an afterimage or who reported less than 5 AIDs during the 8 trials measuring AID were not included in the calculations of the durations, on the grounds that their perception was so incomplete (or their ability to report so poor) that the time values they gave

were of little value. The prior decision had been made to classify these patients as having been diagnosed brain-damaged by the test.

RESULTS

The distribution of total AID scores (the sum of the durations reported for the eight timed trials) is shown in Fig. 1. Only those patients who reported five or more AIDs for the eight timed trials were included in the analysis. The number of cases who were excluded for this reason are listed in Table 1 under the heading "Less than Five Reports." Only 7 of the 123 cases failed to give an AIR in two consecutive pretrials, and only 3 of the remaining 116 cases did not report negative afterimages for at least five of the eight timed trials. The 3 equivocal (undiag-

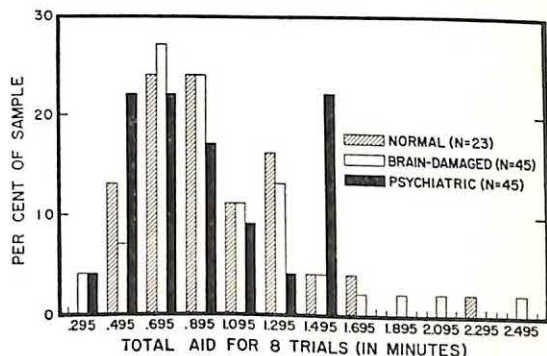


FIG. 1. Distribution of total AID (8 trials) by diagnostic groups.

TABLE 1
MEAN AGE, INTELLIGENCE, AND TOTAL AID (IN MINTUES) FOR ALL DIAGNOSED CASES

Diagnosis	N	Age	IQ	AID	Less than Five Reports
Brain damaged	53	42.78	106.36	.9911	8
Cortical	45	43.15	105.33	.9918	6
Subcortical	8	40.33	113.00	.9867	2
Psychiatric	46	37.85	110.64	.9758	1
Neurotic	34	37.06	110.76	.9309	1
Psychotic	12	40.00	110.33	1.0992	0
Hospitalized Normals	24	50.35	106.78	.9152	1
Total	123	42.35	108.15	.9696	10

nosed) cases are not included in the following analyses since they could not contribute any information to the diagnostic ability of the spiral test.

Table 1 summarizes the mean total AIDs, ages, and IQs for each diagnostic category. An analysis of variance of IQs indicated that the groups did not differ significantly on this variable. The product-moment correlation between total AID and IQ was .034, which was not significantly different from zero. The analysis of variance of age, however, did show that the diagnostic groups differed significantly ($p < .01$) on this variable. The product-moment correlation between total AID and age was found to be $-.092$ ($p > .05$), thereby making a covariance adjustment for age unnecessary.

The analysis of variance of differences in total AID score by diagnostic groups confirmed what was apparent from inspection of Fig. 1. The three diagnostic categories do not differ significantly in the reported duration of the afterimage.

An analysis of variance in which the groups were further divided by the drugs they were taking at the time of testing was performed to see if this added factor might account for the lack of a significant difference between diagnostic groups. Since only Drug Groups 1 and 2 were represented in each of the diagnostic categories, only these two groups were included in the analysis. None of the sources of variation contributed a mean square greater than that for individuals (within

variance) which served as the error term (McNemar, 1955, pp. 304-306).

The lack of differences between diagnostic groups on total AID as a function of drugs might have resulted from too gross a grouping of the various depressant medications. Drug Group 2 was, therefore, further subdivided into the more potent drugs (sparine, thorazine, trilacon, vesprin, and compazine) and the less potent drug, meproamate. Only the psychiatric group was included in this analysis, because there was no certain way to assess the effects of the anticonvulsant medication which provided the main group of drugs prescribed for the brain damaged patients. The latter patients' AID scores were separately analyzed.

The analysis of variance of AID scores for the psychiatric patients, comparing those who received central nervous system depressants with those receiving meproamate and those receiving no centrally acting drug, yielded a nonsignificant F . The AID scores for the 28 brain damaged patients diagnosed as having seizures were compared on the basis of whether or not they were on anticonvulsant medication at the time of testing. Again, the analysis of variance failed to indicate a significant difference attributable to this factor.

The diagnostic groups had differed on yet another variable, order of presentation of the test object; i.e., whether the block or circle was used as a test object for the first four duration trials. Table 2 presents the results of the analysis of variance of order of presen-

tation by diagnostic groups. Again, diagnosis was not significant, while order of presentation of the test object did significantly affect AID. The circle tended to be associated with longer duration times than the block for all diagnostic groups. For those Ss to whom the circle was presented first the mean ratio of their first four to their last four duration times was .76; for those receiving the opposite presentation the mean ratio was .96. Therefore, when the circle was used as a test object on the last four trials, there was a mean decrement of only 4% in AID reported; while, when the block was used on the last four trials, the mean decrement was 24%.

The possibility remained that order of presentation of the test objects might have affected the total AID; and, since the number in each diagnostic category receiving the two orders of presentation were not equally distributed, this might have contributed to the failure of the AID to distinguish between the groups. However, an analysis of variance showed that the total AID did not differ significantly between those presented with the block for the first four trials as against those presented with the circle for the first four trials.

Another factor which might have been of value in distinguishing between the diagnostic groups was stimulation time. A briefer stimulation time might reduce AID more for brain damaged than for other patients. The analysis of variance for each of the stimulation times showed that neither the 15-sec. nor 30-sec. period was successful in significantly differentiating the diagnostic groups.

Stimulation time itself, however, was significantly related to the AID. The mean total

TABLE 3
DISTRIBUTION OF NUMBER OF FAILURES
TO REPORT AFTERIMAGES

Frequency of Failure	Brain Damaged (N=53)	Psychiatric (N=46)	Normal (N=24)
6+	5	1	1
5	0	0	0
4	3	0	0
3	1	1	0
2	1	1	0
1	10	1	2
0	33	42	21

AID following 15-sec. stimulation was .4306 min. and following 30-sec. stimulation was .5388 min. The critical ratio of this difference was significant at less than the .001 level.

Turning from the analysis of AID, the number of zero responses was examined for possible diagnostic significance. The distribution of number of times that patients reported no afterimage is given in Table 3. All cases except the three having no diagnosis are listed. The category "6+" contains those patients who either reported no afterimage for six or more trials determining the duration of the afterimage or who did not report two consecutive afterimages in the pretrials. When the number of zero reports was dichotomized between none and one or more, a chi square of 13.69 was obtained ($p < .01$, $df = 2$). In percentage terms, 37.7% of the brain damaged cases reported one or more failures to perceive the afterimage on either the initial or duration trials. This figure compares with 8.7% of the psychiatric group and 12.5% of the hospitalized normals who failed to report an afterimage one or more times.

DISCUSSION

The results of this investigation fail to support any of the theories which reason that brain damage causes a decrease in cortical efficiency or an increase in cortical inhibition—either of which would result in the reduced duration of the afterimage. The failure of the results to satisfy the predictions of such theories cannot be laid to differences between the criterion groups in age, intelligence, drug status, or gross differences in diagnosis (i.e., psychosis vs. neurosis, cortical vs. subcortical

TABLE 2
ANALYSIS OF VARIANCE OF EFFECT OF ORDER
OF TEST PRESENTATION BY DIAGNOSTIC
GROUPS ON TOTAL AID

Source	df	F
Order	1	9.9396*
Diagnosis	4	.5202
Interaction	4	.0896
Individuals	101	(1,096.77) ^a

* Significant at less than .01 level.

^a Error variance.

brain damage). Neither age nor IQ correlated significantly with AID, nor did depressant drugs have a different effect on AID than no drug when patients with similar diagnoses were compared.

Eysenck's et al. (1957) finding that depressant drugs would foreshorten the AID was not confirmed by this study. However, their experimental Ss were without any known pathology, and the drug's effect on the S's behavior was measured shortly after a single large dose had been administered. The patients in this investigation were selected because they were ill in one way or another, and those who received depressant drugs were given them in doses spaced over a period of time, with the elapsed time between drug ingestion and test administration being uncontrolled. In addition, the depressant drug used by Eysenck et al. (1957) was a barbiturate, whereas in this investigation almost all of the patients in the depressant category were given ataractic or anticonvulsant medication. That these factors are important in assessing the effects of drugs is attested to by Lehmann and Csank (1957) who write:

It is probable that the same drugs would produce differing results after the subjects had taken them regularly for a time because of tolerance or sensitization It is also possible and even probable that severely disturbed patients will show reactions that differ considerably from the pattern observed in healthy control subjects (p. 231).

According to Shapiro's (1954) theory of cortical inhibition, the differences between the diagnostic groups should have been most marked following the longer stimulation period, since cortical inhibition would be aroused more quickly in the brain injured, depriving them of the increase in AID which would follow longer stimulation. Instead, it was found that differences in diagnostic status did not contribute to the significant difference in AID found between the 15-sec. and 30-sec. stimulation periods. The fact that amount of stimulation did affect AID, but not differently for the diagnostic groups, leads to a further question as to the adequacy of cortical inhibition and efficiency theories as now proposed.

The average duration following 15-sec.

stimulation was 6.5 sec. and following the 30-sec. stimulation was 8.0 sec. Considering this brief period of time, the possibility of a sizeable error of measurement caused by individual differences in interpretation and execution of the task is manifest. Some Ss might be expected to delay slightly before reporting cessation of the afterimage, in order to be certain that they were reporting accurately. Others might report cessation even before it had occurred in hopes of bringing the testing to a close more quickly. With the duration period per trial being so small, idiosyncrasies in reporting would have a substantial effect on the results. With this thought in mind, it is interesting to note that the correlation between the AID following the four 15-sec. trials with those following the four 30-sec. trials was .838, indicating a fairly high degree of consistency in reporting the afterimage. It would seem that whatever factors underlie the reporting of the AID they are relatively constant for each individual.

That there was difficulty in interpreting the task is attested to by the fact that the AID associated with the circle as test object was significantly longer than that associated with the block. This finding suggests that two afterimages were involved, one resulting from stimulation by the rotating spiral and one due to the contrast of the black circle on the white ground. It is well established that prolonged exposure to a highly contrasted, stationary stimulus will result in a negative afterimage. It is likely that most of the patients interpreted this afterimage as also being a part of the same phenomenon they were requested to report. The orange block against the neutral ground either did not produce this type of afterimage or did so to a lesser extent.

The conclusion that duration of the afterimage does not differ between brain damaged and other diagnostic groups is consistent with the previously reported study by Page et al. (1957) and inconsistent with that of Spivack and Levine (1957). The former study involved adults, while the latter sampled adolescents. Page et al. unfortunately did not report their data on AID, other than to note that it was not significant. Spivack and Levine, however, report the mean durations for

their brain damaged and control groups, which were 20.2 and 14.8 secs., respectively. Either group has substantially longer AIDs than the 8.0-sec. average for the four 30-sec. stimulation periods of this study. One explanation for this difference can be found in the fact that they seated their Ss five feet from the spiral. At this distance, more rod vision is involved which should *decrease* the AID (Granit, 1928; Wohlgenuth, 1911). At the same time, however, it has been noted (Granit, 1927, 1928) that the duration of an afterimage produced in the periphery of the retina is less distinct and more difficult to measure even by highly trained Ss. It is possible, then, that differences in sampling and in procedure may account for the inconsistency of the results found by Spivack and Levine and those obtained in this study.

One further limitation of this study as well as the two mentioned above is that differences, or lack of differences, in AID found between brain damaged and other groups may be the result of slowed reaction time in the brain damaged. The brain damaged patient may perceive the afterimage as cessating sooner but be slower to report this fact. A method which circumvents verbal report would be necessary to resolve this question. One suggestion, currently being investigated, would be to use electroencephalographic recordings of occipital potentials as the measure of afterimage duration. If the stimulation by a rotating spiral results in changes in the occipital cortex, then it is likely that this change will result in a suppression of occipital alpha for a period following stimulation which is closely related to the perceived duration of the afterimage. The use of such a measure would circumvent verbal report and permit a more direct study of this phenomenon.

Turning to another explanation of these findings, it must be noted that, while AID did not differentiate the diagnostic groups, AIR did significantly separate them. Significantly more brain damaged individuals failed to report a negative afterimage one or more times than did the psychiatric or normal groups. Spitz (1958) has argued that failure to report the illusion is evidence for lack of an "abstract attitude." Certainly there is

considerable ambiguity in the question: What does the line appear to be doing now? It is conceivable that some patients structure this query as referring to the objective condition of the spiral when the experimenter has turned it off. In other cases, failure to report the afterimage may be due to a suspicion on the part of the patient that the perception of the illusion is indicative of some kind of pathology. Thus, denial of the afterimage may be a symptom related to anosognosia, another symptom fairly common in brain damaged individuals (Tobis, Lowenthal, & Maringer, 1957).

Since only 10 of the 123 patients tested did not report a negative afterimage on half or more of the trials, there is no question that the use of a test object increases the likelihood of AIR. That this is true suggests that some of the failures to report the afterimage may be due to confusion about the task. We may look upon this finding as lending further support to the alternative hypothesis that brain damage does not effect the perception of the afterimage, though it may differentially effect the frequency with which this phenomenon is reported.

SUMMARY

An ongoing sample of male veterans admitted to the neurology and psychiatry services of a general hospital were tested with the Archimedes spiral. In addition to determining whether or not the patient could report the negative afterimage, the duration of the negative afterimage was also measured. To decrease the ambiguity of the task, two test objects were used on which the patient was instructed to pass off his afterimage. This technique greatly increased the number of Ss who reported the afterimage. Neither age, intelligence, drugs, nor diagnosis differentiated the patients with regard to the duration of the afterimage. However, when the number of failures to report the afterimage on any of the 10 trials was used, it was found that the brain damaged patients could be separated from the psychiatric and hospitalized normal patients at a highly significant level. It was suggested that these findings call into serious question theories of brain functioning which depend upon a cortical effi-

ciency or cortical inhibition hypothesis. The alternative explanation of studies which indicate that brain damaged patients have greater difficulty in reporting the spiral afterimage seems to be that the brain damaged person is more readily confused by the task. The data from this investigation suggest that brain damaged patients perceive the afterimage as frequently as other patients, but they fail to report their perception more frequently.

REFERENCES

- BERGER, D., EVERSON, R., RUTLEDGE, L., & KOSKOFF, Y. D. The spiral aftereffect in a neurological setting. *J. consult. Psychol.*, 1958, 22, 249-258.
- EYSENCK, H. J., HOLLAND, H., & TROUTON, D. S. Drugs and personality: III. The effects of stimulant and depressant drugs on visual aftereffects. *J. ment. Sci.*, 1957, 103, 650-655.
- GALLESE, A. J. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, 12, 254-258.
- GILBERSTADT, H., SCHEIN, J. D., & ROSEN, A. Further evaluation of the Archimedes spiral aftereffect. *J. consult. Psychol.*, 1958, 22, 243-248.
- GOLDBERG, L. R., & SMITH, P. A. The clinical usefulness of the Archimedes spiral aftereffect. *J. consult. Psychol.*, 1958, 22, 153-157.
- GRANIT, R. Über eine Hemmung der Zapfenfunktion durch Stabchenerregung beim Bewegungsnachbild. *Z. Sinnesphysiol.*, 1927, 58, 95-99.
- GRANIT, R. On inhibition in the aftereffect of seen movement. *Brit. J. Psychol.*, 1928, 19, 147-157.
- KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain-injured. *J. Pers.*, 1952, 21, 118-148.
- LEHMAN, H. E., & CSANK, J. Differential screening of phrenotropic agents in man: Psychophysiologic test data. *J. of clin. exp. Psychopath.*, 1957, 18, 222-235.
- MCNEMAR, Q. in *Psychological statistics*. (2nd ed.) New York: Wiley, 1955.
- PAGE, H. A., RAKITA, N. B. Another application of the determination effect in the determination of accuracy. *Psychol.*, 1957, 21, 159-163.
- PHILBRICK, EMILY B. The value of the test aftereffect as a clinical tool for the test of brain pathology. *J. consult. Psychol.*, 1958, 22, 43.
- PRICE, A. C., & DEABLER, H. L. Diagnostic effect of spiral aftereffect. *J. consult. Psychol.*, 1955, 19, 299-302.
- SAUCER, R. T. The nature of perceptual processes. *Science*, 1953, 117, 556-558.
- SAUCER, R. T., & DEABLER, H. L. Perception of apparent motion in organics and schizophrenics. *J. consult. Psychol.*, 1956, 20, 385-389.
- SHAPIRO, M. B. A preliminary investigation of the effects of continuous stimulation on the perception of "apparent motion." *Brit. J. Psychol.*, 1954, 45, 58-67.
- SPIZ, H. H. The present status of the Kohler-Wallach theory of satiation. *Psychol. Bull.*, 1958, 55, 1-28.
- SPIVACK, G., & LEVINE, M. The spiral aftereffect and reversible figures as measures of brain damage and memory. *J. Pers.*, 1957, 25, 1-11.
- TOBIS, J. S., LOWENTHAL, M., & MARINGER, S. Evaluation and management of the brain-damaged patient. *J. Amer. Med. Ass.*, 1955, 65, 2035-2041.
- WOHLGEMUTH, A. On the after-effect of seen movement. *Brit. J. Psychol., Monogr. Suppl.*, 1911, 1, 1-117.

(Received July 29, 1959)

A STUDY OF IDENTIFICATION IN MALE HOMOSEXUALS¹

JUDY CHANG AND JACK BLOCK

University of California, Berkeley

In discussing the etiology of male homosexuality, Freud (1952) has written:

The typical process . . . [in homosexuality] is that a few years after the termination of puberty, the young man, who until this time has been strongly fixated to his mother, turns in his course, identifies himself with his mother, and looks about for love-objects in whom he can re-discover himself, and whom he wishes to love as his mother loved him (p. 240).

According to Freudian theory, therefore, homosexuality in males is a manifestation of an identification with the mother figure rather than the usual identification with a father figure. As a consequence of his twisted identification, the homosexual seeks narcissistic rather than anaclitic love-objects.

The purpose of the present study is to test a consequence of this notion, namely, that male homosexuality involves both a turning *toward* the mother as a figure with which to identify and the turning *away* from the father as an identification figure. Specifically, the major hypothesis of this investigation was that frankly homosexual males have a pattern of parental identification such that they are more strongly identified with their mothers and, conjointly, more strongly disidentified with their fathers than are males in a relevant contrast group. This conjoint hypothesis of course can be further partitioned into two separate hypotheses of interest, permitting comparison of the homosexual and nonhomosexual subjects (Ss) with regard to their *absolute* degree of identification with mother and with father figures. The two subhypotheses that follow from the major hypothesis are: (a) overt homosexual males, when compared to nonhomosexuals, will show a stronger iden-

tification with their mothers, and (b) overt homosexual males will show less identification with their fathers than will members of the control group.

The meaning of the Freudian term identification has been the subject of some controversy (Sanford, 1955; Stoke, 1950). For research purposes, various operational translations of the concept have been employed. The operational meaning of identification adopted for the purposes of the present study is based upon the following statement by Freud (1933): ". . . when a boy identifies with his father he wants to *be like* his father" (p. 90). According to Freud, at least in this late statement by him on the matter, identification with an object is said to have occurred when an individual strives to take on attributes which are those of that object as the individual perceives that object; this object is no more and no less than one's ego ideal—the object ". . . towards which it . . . [the ego] strives" (Freud, 1933, p. 92). In this study, then, a measure of the extent of identification with one's mother or father following directly from this conceptual definition was derived by finding the degree of correspondence between the attributes of one's ideal self, i.e., the person one would like to be, and the attributes of one's mother or father, as described by the identifier. The higher the degree of correspondence between one's ego ideal and one's parent, the greater the identification with that parent.

METHOD

Two groups of Ss were used. One group, the experimental group, consisted of 20 overt homosexual males. These Ss were engaging in homosexual practices, i.e., most of them were involved in homosexual marriages. None of these Ss were receiving or, with the exception of one, had ever received psychotherapy. Thus, all of these Ss appeared to

¹ This investigation, part of a larger research project, was supported by research grant M-1078 from the National Institute of Mental Health of the United States Public Health Service.

have accepted their homosexuality, a most important and perhaps unusual property of this sample. Their cooperation to serve as Ss was obtained through personal and necessarily confidential channels. These Ss were aware that they had been asked to participate in the experiment because of their homosexual practices but had not been informed of the purpose of the experiment. A second group of 20 males from the same urban community and equivalent with respect to age and education was used as a contrast group. Eight of these Ss were married, and twelve were still attending college. The mean age of the homosexual males was 31.9 years with a range from 22 to 39, that of the contrast group, 29.7 years with a range from 22 to 39. The mean educational level was 14.5 years for the homosexual group and 14.9 years for the contrast group.

The measure of identification with parent was obtained through the use of a list of 79 adjectives selected to permit a comprehensive personality description.² Each S was instructed to describe his "ideal self" by the use of the adjectives, marking an X if he felt that the adjective was particularly characteristic of his ideal self or an O if it was particularly uncharacteristic. The S was allowed to make only and exactly 30 X's and 30 O's in order to avoid the confusions engendered by response sets (Cronbach, 1950). Following this description, S was required to describe, in the same manner, "your mother," "your father," and "yourself (as you see yourself)." In order to facilitate independence of the several descriptions, each object description was made on a separate page of the experimental booklet.

Of 35 booklets circulated into the available homosexual population, only 22 were returned. Of these, only 20 were found to be usable in the sense of being complete or correctly filled out. Data for the contrast group then was collected by distributing booklets among individuals who were similar to the homosexual group with respect to age and educational level. Of 31 individuals requested to participate in the experiment, only 21 were willing to do so. One of these 21 control Ss was arbitrarily eliminated from the subsequent data analysis in order to equalize the sample size. In view of the fractional returns from the populations sampled, the possibility of some kind of selective bias in the results should, of course, be recognized.

For each S, a mother-identification score and a father-identification score was obtained by comparing his description of his ideal self with that of his mother and with that of his father, respectively. For a given adjective, a score of one was given if it was marked X, O, or left unmarked on both the ideal-self- and parent descriptions. An identification score was obtained by summing the number of these correspondences between the ideal-self-description and the parent description. The separate mother- and father-identification scores thus derived were used to test the two subhypotheses. For the major hypothesis, that male homosexuality represents simultaneously an idealization of mother together with

a disparagement of father, a derived score was obtained by subtracting for each S his father-identification score from his mother-identification score. The score distributions for the mother-identification and father-identification variables had quite equivalent means and standard deviations (40.7 and 11.3; 41.0 and 11.5, respectively). Accordingly, it was legitimate to calculate the derived score from the original data distributions without going to the trouble of standardizing the separate score distributions. Thus, this derived score would be high if an S's mother-identification score was high and/or his father-identification score was low. In addition, a self-acceptance score was obtained by comparing the ideal-self-description with the self-description, scoring for correspondence in the manner described above. These four scores provided the basic data for analysis.

For each comparison of interest, the Mann-Whitney Test (1947) was used to calculate the significance of the difference between the scores obtained by the experimental and control groups. With the exception of the analysis of self-acceptance scores, all the statistical results reported in this study are based on one-tailed tests of significance because the direction of difference was predicted in advance.

RESULTS

For the homosexual group, the derived scores expressing the primary variable, i.e., distance between mother and father on S's identification continuum, are significantly greater (.001) than the derived scores of the contrast group. Some indication of the separation between the two groups on this variable is provided by the finding that 14 of the 20 homosexuals have positive scores where only 4 of the 20 nonhomosexuals are on the positive side of zero. The primary hypothesis, that overtly homosexual males are relatively more strongly identified with their mothers than with their fathers, is therefore supported by this finding.

When the primary variable is partitioned into its components (the mother-identification score and the father-identification score) and separately analyzed, a better perspective is gained on the nature of the preceding finding. When a comparison of the mother-identification scores of the two groups is made, the homosexual Ss show a significantly greater degree of identification with their mothers. The difference between the sums of ranks of mother-identification scores of the two groups of Ss is significant at the .03 level. An analysis of the father-identification scores of the two groups of Ss shows a not quite

² A copy of this list of adjectives is available by writing to the authors.

significant but still appreciable trend in the direction of less identification with father by the homosexual group. The one-tailed test results in a p value of .07, almost reaching the conventional level of significance. From these separate and rather equivalent findings, it is clear that the striking separation between the two groups in the test of the first hypothesis is a *joint* reflection of the two constituent scores rather than being dominated by but one of these. This kind of result is both a subtle and a more interesting one.

Self-Acceptance

The homosexual and control groups did not differ significantly in their extent of self-acceptance, where self-acceptance is defined as the degree of correspondence between the perceived self and ideal self. It should be especially noted that there is no perceptible trend for the contrast group to be relatively more self-accepting, as indicated by the two-tailed p value of .64 for the obtained difference between the sums of ranks. This finding is a rather interesting one for it calls attention to the special nature of this homosexual sample as one *not* psychiatrically disturbed and apparently well-accepting of their homosexual activities.

Analysis of Ideal-Self-Descriptions

The results in the preceding analyses might have been due to differences between the kinds of ego ideals of the homosexual and control groups. For this reason, an item analysis of the adjective descriptions of the ideal selves of the two groups of Ss was done. Only 2 of the 79 adjectives discriminated (at the .05 level) between the ideal-self-descriptions of the two groups. The conclusion is made, therefore, that the differential valuations by the homosexual Ss and control Ss of their parents are *not* due to different ego ideals. For the sake of completeness, it may be recorded that nonhomosexual Ss considered "sympathetic" as an attribute of their ideal self more frequently than did homosexual Ss. Homosexual Ss, more than control Ss, tended to deny the characteristic "dependent" as an attribute of the ideal self. No additional adjectives were found to discriminate when the threshold of significance was lowered to the .10 level.

Analysis of Self-Descriptions

Since the homosexual and control Ss did not differ in their degree of self-acceptance and in their ideal-self-descriptions, one might expect that they would also not differ in their self-perceptions. But if this expectation were to be realized, then some of our preceding findings become questionable. The reasoning behind this remark is that homosexuals and normals are quite obviously rather different sorts of people and, consequently, these differences should, if only on a priori grounds, be manifest if a reasonably sensitive instrument is employed. Failure to find these differences would be presumptive evidence of the technique's inadequacy, a conclusion that might also explain away the failure to find differences with respect to self-acceptance and ego ideals. On the other hand, it is possible, despite the failure of the self-acceptance and ideal-self-analyses to show differences, for the self-description analyses to separate the groups. This possibility arises because the self-acceptance for neither group is so high as to require (because of equivalent ideal selves) that the content of the self-descriptions be equivalent. With moderate correlations of self- with ideal-self-descriptions, the different dimensions of self-description are free to operate and, hence, be seen.

Item analysis of the self-descriptions of homosexual and control Ss finds six adjectives discriminating at the .05 level, with an additional six adjectives meeting the .10 criterion—results not accountable by a proper statement of chance significance.³ Moreover, the discriminating adjectives are consonant with the concept of the homosexual male. Compared to control Ss, the homosexual male sees himself as "affected, dependent, determined, personally charming, restless, and *not* tact-

³ By reference to "a proper statement of chance significance," we are alluding to a recently developed recognition that conventional expectations of chance significance have been appreciably wrong rather often. In the situation where the response distributions of both groups being compared are highly skewed (or in the dichotomous case, are extremely unbalanced as in the present study), at the, for example, .05 level, many less than 5% of the findings will be significant when chance alone is operating. The logic underlying this assertion and some empirical findings are detailed in a manuscript available from the second author.

less" (all of these significant at the .05 level). The picture can be extended by noting that the homosexual tends to see himself as "impulsive" but not "frank" or "easily embarrassed" or "ambitious." Our homosexual males are less often "friendless or introspective or self-aware" than the control Ss, all of these last differences reaching the .10 level.

Although the discriminating characteristics cluster and portray someone with the personality traits usually associated with homosexuality in this culture, it is especially noteworthy that such adjectives as "confident, psychologically secure, relaxed, dissatisfied, helpless, worried" and the like failed to discriminate between the self-perceptions of the two groups.

DISCUSSION

The major hypothesis of this study, that frankly homosexual males more than control males show a positive identification with their mothers together with a disidentification with their fathers, has been supported by the results of this study. Subsidiary hypotheses have also been supported. What now shall we make of this finding?

Clearly, no one study is able to test the Freudian hypothesis, in all its complexities, about homosexuality. The present research does, however, in its limited way, support the psychoanalytic position. Additional studies, by the present and alternative methods, are required to investigate the multifaceted Freudian theory that the relatively strong identification with mother found in homosexual males has been preceded by and is the result of a strong fixation upon the mother, that the starting point of homosexuality comes quite discretely in adolescence, and that homosexuality represents a narcissistic object-choice. Such a research represents an undertaking of a much larger dimension than the present one.

With respect to the present study, the unusual nature of the experimental sample used—self-accepting homosexuals—should especially be noted. Our results may well be an expression of the special character of this group. In more conflicted homosexuals, where identifications have been less decisively made, the presently obtained relationships can be expected to exist less strongly. One should

find, among individuals with more ambiguity of identification, that there is less self-acceptance and more personal despair than was observed among the effectively functioning and well-received homosexuals studied in the present research.

SUMMARY

The present study was concerned with a test of the Freudian notion that homosexuality in the male is based upon an over-identification with the mother figure together with an under-identification with the father figure. Two groups of Ss, equivalent with respect to age, education, and socioeconomic level, were used: a group of 20 males engaging, apparently successfully, in homosexual practices, and a group of nonhomosexual males. Each S described his ideal self, his mother, his father, and himself, by the use of a list of adjectives. On the basis of an "ego ideal" conception of identification, various identification scores and a self-acceptance score were obtained for each S, by comparing his description of his ideal self with that of the other individuals he described.

As tested, the hypothesis was supported for it was found that overt homosexual males were more identified with their mothers and more disidentified with their fathers than were the nonhomosexual males with whom they were compared. In addition, it was found that the two groups of Ss did not differ significantly in their degree of self-acceptance nor in regard to the kind of ego ideal toward which they aspired. The findings were briefly discussed.

REFERENCES

- CRONBACH, L. J. Further evidence on response-sets and test design. *Educ. psychol. Measmt.*, 1950, 10, 3-31.
- FREUD, S. *New introductory lectures on psychoanalysis*. New York: Norton, 1933.
- FREUD, S. *Collected papers*. Vol. 2. London: Hogarth, 1952.
- MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, 18, 50-60.
- SANFORD, N. The dynamics of identification. *Psychol. Rev.*, 1955, 62, 106-118.
- STOKE, S. M. An inquiry into the concept of identification. *J. genet. Psychol.*, 1950, 76, 163-189.

(Received July 29, 1959)

THE AFTEREFFECT OF SEEN MOVEMENT AND BRAIN DAMAGE

R. H. DAY

University of Sydney

If a moving stimulus object which has been visually fixated for about one minute is stopped suddenly, there occurs a characteristic illusory movement in a direction opposite to that of the real movement. This aftereffect of seen movement is experienced following fixation of various types of moving stimulus objects including a rotating spiral pattern, a sectorized disc, or a pattern of black and white bands (waterfall illusion). The phenomenon is not restricted to the stimulus object but can be observed in the field of the closed eye and on other textured surfaces. If one eye is used during fixation and that eye closed on the cessation of real movement, the aftereffect can be observed using the non-stimulated eye.¹

Recent studies have drawn attention to a relationship between this aftereffect and brain damage. So far only the rotating spiral pattern has been used as a stimulus object (Aaronson, 1958; Davids, Goldberg, & Laufer, 1957; Freeman & Josey, 1949; Gallese, 1956; Garrett, Price, & Deabler, 1957; Holland & Beech, 1958; Price & Deabler, 1955; Spivack & Levine, 1957). Results from these studies have indicated that organically impaired subjects (Ss) can be discriminated from normal or functionally disordered controls with a high degree of certainty. Generally it has been found that, whereas normal individuals experience a vivid and unmistakable aftereffect, cases of brain damage either report no aftereffect or one which is reduced in its duration or vividness. Some studies, however, have not confirmed this and have reported that little or no differentiation can

be made between brain damage and normal controls (Goldberg & Smith, 1958; Johnson, Bauer, & Brown, 1959; Page, Rakita, Kaplan, & Smith, 1957; Philbrick, 1959; Pickersgill & Jeeves, 1958).

It is intended here to direct attention to issues which appear to have been overlooked and which have an important bearing on research in this area. It is believed that failure to consider certain behavioral outcomes of brain damage as well as insufficient attention to methodology and sampling have resulted in the lack of agreement between the results from independent studies.

VISUAL FIXATION AND MOVEMENT AFTEREFFECTS

The aftereffect of seen movement occurs when the eye is fixated on a stationary point during the period of stimulation by the moving pattern (Holland & Beech, 1958; Wohlge-muth, 1911). If the eye follows the moving pattern, little or no aftereffect occurs. In order to confirm this 12 normal Ss were required to follow one of the sectors on a black and white sectorized disc on one trial and to fixate the center of the disc on another. During the following trial eight Ss failed to report an aftereffect and four reported a slight but brief aftereffect. All Ss experienced a strong aftereffect (apparent movement in the reverse direction) following fixation of the center of the disc.

When the eye follows a point on a moving stimulus pattern, the retinal image of the pattern either remains stationary or moves less relative to the retina. When a fixed point such as the center of the disc is fixated, in retinal image moves with respect to the retina. An essential condition, therefore, for the occurrence of a clear aftereffect of seen movement

¹The apparent interocular transfer of an aftereffect from the stimulated to the nonstimulated eye cannot be interpreted as evidence for its central (cortical) locus (Day, 1958).

is motion of the retinal image of the stimulus pattern. This condition is achieved by steady fixation of a stationary point in or near the stimulus object.

FIXATION AND BRAIN DAMAGE

In an investigation of the cortical integration of ocular movements it has been shown that, whereas some organics experienced difficulty in voluntarily changing the point of fixation, their eyes followed a moving object involuntarily (Holmes, 1938). In an account of the effects of frontal brain damage on eye movement and fixation Holmes has stated:

This condition at first shows all the appearance of loss of paralysis of movement, but that this is not present becomes obvious when the subject is asked to follow with his eyes a slowly moving object; then, providing the rate of movement is uniform and not too fast a full range of deviation may be obtained . . . when the patient looks at a rotating drum on which lines or figures are displayed the eyes also deviate in the direction in which the figures move . . . We therefore find that though voluntary effort fails to produce movement the eyes follow a moving point the images of which fall on the retina (p. 106).

In discussing the work of Holmes (1938) Sherrington (1940) has presented a similar view:

But if asked to follow with his eyes a slowly moving object, his eyes do so. That is, though willed effort fails to produce the act, the eyes follow 'of themselves' a moving point whose image is fixated by them. The eyes remain anchored to that object by the 'fixation reflex.' Those afflicted with this condition when they want to look at something fresh, close the eyes for a moment, or toss the head, or place a hand between the eyes and the object fixing their gaze. By breaking the gaze a moment they free themselves from the fixation reflex; though by willed effort they cannot break it (p. 226).

Thus in those cases in which voluntary eye movements and visual fixation have become difficult or unattainable, but in which optical fixation reflexes are preserved, the eyes remain fixated on the moving object. This symptom of frontal brain damage has been called "following movement" by Walsh (1957). It would be expected then that the moving elements of a stimulus pattern would tend to be followed, rather than a stationary point fixated. It has been pointed out that

this method of observation does not result in a persistent or vivid aftereffect. It seems probable that the absence, weakness, or short duration of the aftereffect of seen movement observed in certain cases of brain damage derives in some measure from an inability to meet the conditions of visual fixation necessary for its occurrence. Even if fixation of a stationary point is achieved during part of the period of stimulation, the aftereffect would be reduced as compared with that resulting from steady fixation during the full period of stimulation. The length of the period of stimulation is an important determinant of the aftereffect (Wohlgemuth, 1911).

A possible relationship between the well known distractability of organically impaired Ss and reduced aftereffects of seen movement has already been pointed out (Holland & Beech, 1958). It has been suggested that such distractability could result in inadequate fixation and thus contribute to the weakness of the aftereffect. The view proposed here, however, is that the weakness of the aftereffect is based on an inability on the part of the organic S to fixate rather than on a tendency to inattention.

MOVEMENT AFTEREFFECT AND THE LOCUS OF BRAIN DAMAGE

The control of eye movements is not confined to a single cortical center. Stimulation of the frontal region (Brodman's Area 8), the inner aspects of the occipital lobes and temporoparietal centers produces eye movements. Penfield and Rasmussen (1950) located an eye movement (turning) center at the junction of the "face" and "hand" areas in the sensorimotor sequence but extended anteriorly with respect to the rest of the sensorimotor strip. Eye turning has also been elicited in some cases by stimulation of the temporal and parietal lobes (Penfield & Boldrey, 1937).

The most marked disturbances of voluntary fixation occur when the frontal region is affected, especially when there is involvement of the frontal projection fibers in the internal capsule. Disturbances of fixation also occur when the fibers in the anterior region of the midbrain are affected (Holmes, 1938).

It would be expected, therefore, that the degree to which the aftereffect of seen movement is affected would depend upon the nature, locus, and extent of brain damage, the most marked deterioration occurring with damage to those centers controlling voluntary fixation. There have been no studies so far in which the locus and magnitude of damage have been considered in relation to the aftereffect. The most common practice has been to compare the aftereffect reported by a heterogeneous organic group with that reported by normal and functionally disordered controls. Since the constitution of the brain damaged groups has varied widely between the numerous investigations, it is not surprising that inconsistencies between results have occurred.

It seems probable that the discrepancies between independently collected data is due in part to a failure to consider the locus, chronicity, and extent of brain damage. Considering the wide variety of behavioral symptoms manifested by such diverse conditions as epilepsy, traumatic brain injury, Parkinson's disease, and postsurgical conditions there is little justification for regarding brain damage as a single category characterized by a common experience (or lack of it) following visual stimulation. It is suggested here that the widely differing results derive from the varying effects of different types of brain damage on ability to fixate under the conditions employed.

SCORING METHODS

Two methods for scoring the aftereffect of seen movement have been used. The simplest method is that of scoring the reported occurrence and nonoccurrence of the aftereffect 0 and 1, respectively, and an equivocal report 0.5 (Freeman & Josey, 1949).² A variant of this method is to score nonoccurrence 0, occurrence 1, and to score 0.5 when a single phenomenal feature only (contraction-expansion, change in size or approach-recession) is reported (Garrett et al., 1957; Philbrick,

1959; Price & Deabler, 1955). A second method requires the S to report the duration of the aftereffect (Holland & Beech, 1958; Page et al., 1957; Spivack & Levine, 1957).

It has been found that, whereas some brain damaged Ss fail to report the aftereffect spontaneously, they have reported it when the nature of the phenomenon has been described to them. This has been most noticeable when a rotating black and white sectorized disc has been used as the stimulus object. The aftereffect from this pattern (apparent reversal of movement) is much less vivid than from a spiral pattern. It has been observed also that organically impaired Ss often report a single feature spontaneously but will describe another feature following a description of it.

It is conceivable that with respect to both the presence or absence of the aftereffect and the number of reported features the differences between organic and normal Ss derive from the adequacy with which they report. If the duration or persistence of the aftereffect were used as the score after it has been established that the S experiences the phenomenon, the reliability of the scores could be readily established. It has been found that duration can be reported with confidence by most organic Ss with the exception of those who exhibit marked deterioration and confusion.

DISCUSSION

Two explanations of the absence or weakness of the aftereffect of seen movement reported by organic Ss have been suggested. The first postulates a synthesizing or integrating process in the CNS responsible for movement perception. It is argued that this process is either absent or adversely affected by brain damage (Price & Deabler, 1955). This view is based upon a recent theory of motion perception (Saucer, 1953, 1954). A second explanation derives from a concept of central nervous inhibition. The effect of brain damage is to intensify cortical inhibitory processes and thus to weaken the aftereffect (Holland & Beech, 1958; Price & Deabler, 1955). Each of these explanations emphasizes the role of central nervous mechanisms as determinants of the aftereffect of seen movement.

²Whereas Freeman and Josey (1949) score nonoccurrence of the aftereffect as 1 and occurrence 0, other investigators have scored 1 for occurrence and 0 for nonoccurrence. The latter would appear to be a more meaningful method.

The view expressed here is that the failure of some organic Ss to experience the aftereffect from a moving pattern, or to experience it with less vividness than normal Ss, is due to a tendency to visually follow the elements of a moving pattern. This tendency is consequent upon damage to those parts of the brain initiating and controlling voluntary fixation and a replacement of these voluntary activities by reflex movements controlled from occipital centers. Although visual observation of the eye movements of brain damaged and normal Ss has shown that the former exhibit markedly more frequent and grosser eye movements during "fixation," this finding has yet to be confirmed by ophthalmographic recording.

In view of the wide variations in the locus, extent, and nature of brain damage there seems little justification for attempting to establish a relationship between a sensory aftereffect and brain damage. It is possible that this simple test may prove useful in differentiating between types of damage rather than differentiating it from other conditions. If the hypothesis outlined here is supported by subsequent data, a more efficient technique would be that of observing the magnitude and direction of eye movement during the period when S is attempting to fixate. Such a method would have the advantage of dispensing with subjective reports of an experience which may be an indirect outcome of brain damage.

SUMMARY

Recent studies have shown that with brain damaged Ss the aftereffect of seen movement from a rotating spiral pattern is either absent or weaker than that experienced by normal Ss. Some studies have failed to confirm this observation.

A necessary condition for the occurrence of this aftereffect is steady fixation of a stationary point in or near the moving stimulus pattern so that the retinal image of the pattern moves relative to the retina. Voluntary fixation and eye movement is affected in certain cases of brain damage so that individual's eyes tend to follow the moving pattern. It is suggested that the weakness or

absence of the aftereffect from a moving pattern derives from this failure of voluntary fixation.

Since the main centers controlling voluntary eye movement and fixation are located in the frontal region of the brain, it would be expected that damage affecting that region would have the most marked effect on the aftereffect.

Inconsistencies between the results from various studies probably derive in large part from a failure to consider the locus, extent, and nature of brain damage as well as the inadequacies of the methods used to measure the aftereffect.

Some preliminary results in support of the eye movement hypothesis have been presented.

REFERENCES

- AARONSON, B. S. Age intelligence, aphasia and the spiral aftereffect in an epileptic population. *J. clin. Psychol.*, 1958, 14, 18-21.
- DAVIDS, A., GOLDBERG, L., & LAUFER, M. W. The relation of the Archimedes spiral after-effect and the Trail Making Test to brain damage in children. *J. consult. Psychol.*, 1957, 21, 429-433.
- DAY, R. H. On interocular transfer and the central origin of visual after-effect. *Amer. J. Psychol.*, 1958, 71, 784-789.
- FREEMAN, E., & JOSEY, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, 62, 794-797.
- GALLESE, A. J. Spiral after-effect as a test of organic brain damage. *J. clin. Psychol.*, 1956, 12, 254-258.
- GARRETT, E. S., PRICE, A. C., & DEABLER, H. L. Diagnostic testing for cortical brain impairment. *Arch. Neurol. Psychiat.*, 1957, 77, 223-225.
- GOLDBERG, L. R., & SMITH, P. A. The clinical usefulness of the Archimedes spiral in the diagnosis of organic brain damage. *J. consult. Psychol.*, 1958, 22, 153-157.
- HOLLAND, H. C., & BEECH, H. R. The spiral after-effect as a test of brain damage. *J. ment. Sci.*, 1958, 104, 466-471.
- HOLMES, G. The cerebral integration of ocular movements. *Brit. med. J.*, 1938, 2, 107-112.
- JOHNSON, D. E., BAUER, R. W., & BROWN, D. R. Visual afterphenomena in diagnosis. *J. consult. Psychol.*, 1959, 23, 90.
- PAGE, H. A., RAKITA, G., KAPLAN, H. K., & SMITH, N. B. Another application of the spiral aftereffect in the determination of brain damage. *J. consult. Psychol.*, 1957, 21, 89-91.
- PENFIELD, W., & BOLDREY, E. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, 1937, 60, 389-443.

- PENFIELD, W., & RASMUSSEN, T. *The cerebral cortex of man*. New York: Macmillan, 1950.
- PHILBRICK, E. B. The validity of the spiral after-effect as a clinical tool for diagnosis of brain pathology. *J. consult. Psychol.*, 1959, 23, 39-43.
- PICKERSGILL, M. J., & JEEVES, M. A. Aftereffect of movement produced by rotating spiral. *Nature*, 1958, 182, 1820.
- PRICE, A. C., & DEABLER, H. L. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, 19, 299-302.
- SAUCER, R. T. Nature of perceptual processes. *Science*, 1953, 117, 556-558.
- SAUCER, R. T. Processes of motion perception. *Science*, 1954, 120, 806.
- SHERRINGTON, C. S. *Man on his nature*. Cambridge: Cambridge Univer. Press, 1940.
- SPIVACK, G., & LEVINE, M. The spiral after-effect and reversible figures as measures of brain damage and memory. *J. Pers.*, 1957, 25, 1-11.
- WALSH, E. G. *Physiology of the nervous system*. London: Longmans Green, 1957.
- WOHLGEMUTH, A. On the after-effect of seen movement. *Brit. J. Psychol. Monogr. Suppl.*, 1911, No. I.

(Received July 29, 1959)

PSYCHOTHERAPISTS' CLINICAL JUDGMENTS AND ATTITUDES TOWARDS PATIENTS¹

MARTIN S. WALLACH AND HANS H. STRUPP
School of Medicine, University of North Carolina

Theoretically, and from the standpoint of research, the therapist appears to be finding his way back into therapy. This trend is in keeping with a renewed interest in countertransference or, more broadly, the therapist's contribution to the treatment process. Mensh (1956) states, "... study of the therapist is now more explicitly indicated, an emphasis which has appeared with trends toward introducing experimental design qualities into research in psychotherapeutic process" (p. 342). Butler (1952) asserts that "... exact observation of the therapist is a necessary precondition to understanding the behavior of the client" (p. 378); similarly, Gitelson (1952), Jackson (1956), and Wolberg (1954) stress the importance of therapist variables. Perhaps the primary impetus for focusing upon the therapist stems from a belief that orthodoxy is waning and "Viewing therapy as a very intimate relationship and recognizing a trend toward less restrictive therapist roles, the expectation arises that differential therapist characteristics would have systematic effects on process and outcome" (Winder, 1957, p. 316).

Previous attempts to explore the relationships between therapist variables and therapist performance in treatment have considered mainly those factors which most readily lend themselves to objective quantification: i.e., the therapist's professional affiliation, theoretical allegiance, length of experience, number of hours of personal analysis, etc.

¹ This paper is a summary of a dissertation submitted in partial fulfillment of the PhD degree to the Department of Psychology, University of North Carolina.

This research is part of a larger project supported by a research grant (M-2171) from the National Institute of Mental Health, Public Health Service.

The potential importance of less tangible personality variables, long recognized in Freud's conception of countertransference, is beginning to attract the attention of investigators seeking objective documentation. For instance, Parloff (1956) found that the therapist relates most satisfactorily to those patients who most clearly approximate his "Ideal Patient." Strupp (1958) has shown that the therapist's attitude toward the patient was significantly correlated with the form of therapy, frequency and length of treatment, the kinds of technical problems expected in therapy, and certain qualitative aspects of the techniques the therapist expected to employ. These observations strongly support the notion that emotional and attitudinal factors in the therapist may have an important bearing upon his clinical evaluations.

The therapist, as a function of his life experiences, approaches each initial interview with needs, expectations, and wishes of his own. It may be hypothesized that, if his expectations are sufficiently realized, he will consider the situation as "rewarding," and a "warm" attitude toward the patient is likely to develop. More specifically, if in an initial interview the prospective patient approximates the therapist's conception of an "ideal patient," he may develop a warm attitude toward the patient. While the conception of "ideal patient" may differ from therapist to therapist, in a general sense, it may be assumed to imply a certain congruence between the kind of help the patient is seeking and the kind of help the therapist is able to provide, as well as an explicit or implicit belief on the patient's part that the treatment will be effective. Patients exhibiting such hopes

and beliefs may be said to be "motivated for therapy" or "ready for a psychotherapeutic experience." Thus, patients whose beliefs, desires, and expectations are similar—in some respects at least—to corresponding expectations held by the therapist may be said to arouse a warmer attitude in the therapist. This attitude, it is hypothesized, will be sufficiently strong to give rise to a kind of "halo effect" overshadowing certain "reality factors." Operationally, this may be evidenced by the character of the therapist's perception of the patient, treatment planning, and the like.

The initial halo effect, it is further hypothesized, will be self-realizing. This is to say, having achieved a "comfortable" or "warm" feeling toward the patient, the therapist will continue to focus on perceptions which are likely to engender further this effect, and he will propose a therapeutic regime which will reflect this attitude. Moreover, he may be more inclined to view the patient as a "good patient" and work toward a closer therapeutic contact with the patient. The present investigation aims to explore the relationships between the therapist's attitude toward the patient on the one hand and his perceptions of the patient and treatment plans on the other. The causal direction, of course, remains an open question.

HYPOTHESES

1. A patient's verbal or nonverbal expectations concerning his "motivation for therapy" or "readiness for therapy" is an important determinant of the therapist's attitude toward the patient. In particular, the more highly motivated patient will engender in the therapist a warmer attitude.

2. The therapist's attitude toward the patient, whether a function of the patient's motivation for therapy or other factors, is correlated with the therapist's perception of, and treatment plans for, the patient. More specifically, a warmer therapist attitude toward the patient is associated with more favorable perceptions of the patient, including clinical judgments, prognostic estimates, and treatment plans implying closer therapeutic contact.

PROCEDURE

To test these hypotheses it was planned to present a group of therapists with a patient manifesting relatively high motivation for therapy, and a comparable group of therapists with the same patient showing relatively low motivation for therapy. Furthermore, to lend greater generalizability to the findings, therapists were presented with two patients.

Two written case histories of neurotic patients were adapted for experimental purposes. Each case history consisted of alternate forms which were identical in all respects, except that one form presented the patient as highly motivated and the second form as poorly motivated for therapy. "Motivation" was communicated primarily in terms of the patient's eagerness to seek and accept psychotherapeutic help. Each case history was followed by an identical questionnaire of 27 items eliciting the respondents' clinical impressions, treatment plans, and attitudes toward the patient. Subjects were presented with the case histories of both patients, the degree of motivation being varied systematically. In addition, several biographical questions were appended to the second questionnaire.

Sample

Data were collected from medical psychotherapists actively engaged in the practice of psychotherapy. Three sources were utilized: (a) physicians known to the investigators personally, (b) colleagues of psychologists at selected installations, and (c) psychiatrists selected at random from the Directory of the American Psychiatric Association. Of

TABLE 1
BIOGRAPHICAL DATA OF RESPONDENTS

	<i>N</i>	Male	Female	Mean Age	Mean Hours of Personal Analysis	Mean Years of Experience
Psychiatric Residents	11	10	1	33.91	140 (2) ^a	3.05
Psychiatrists	37	34	3	35.86	487 (26)	7.27
Psychoanalysts	34	30	4	44.47	625 (34)	14.31
Totals	82	74	8	39.17	549 (62)	9.62

^a Figures in parentheses indicate number of respondents who had personal analysis.

300 sets of material distributed personally or by mail, 82 were included in the analysis. One hundred and eighty-four sets were never returned, 26 respondents asked to be excused, and 8 respondents completed and returned the questionnaires after the analysis of the data was well in progress. Analysis of the biographical data (age, sex, professional level, years of experience, number of hours of personal analysis) supplied by respondents indicated that the sample was drawn from a relatively homogeneous universe of experienced medical psychotherapists (see Table 1), and that in terms of these variables the various subsamples showed no significant variations.

RESULTS

Position Effects

As a first step in the statistical analysis it seemed important to investigate the possible influence of order of presentation (patient and degree of motivation) upon questionnaire responses. Since 24 multiple-choice items were answered under four experimental conditions, 96 comparisons were called for. Chi squares and *t* tests of these comparisons yielded 15 results significant at beyond the .1 level. These significant position effects appeared to be largely due to the fact that one of the patients seemed objectively more disturbed, so that respondents studying the male (more disturbed) patient first tended to attribute greater personality strengths and assets to the female (less disturbed) patient when her case was presented second.

Effects of Communicated Patient Motivation

Chi square or *t* tests were employed to ascertain effects of the patient's motivation, as communicated in the case histories, upon responses to the 24 multiple-choice items. Five figures significant at or beyond the .05 level were obtained for the male case, and nine figures significant at this level were obtained for the female case.

When Mr. J. was presented as possessing high motivation for therapy, respondents tended to attribute to him greater ego strength, more anxiety, more insight into his problems, greater motivation for therapy (check item), and they expressed greater willingness to accept the patient for treatment. Contrary to the hypothesis, therapists' attitudes toward the patient did not vary significantly. However, there was a (statistically

nonsignificant) trend in the predicted direction—a slight tendency for therapists to claim a warmer attitude toward the patient when he was presented as being more highly motivated for therapy.

When Mrs. D. was presented as possessing high motivation for therapy, respondents estimated ego strength, insight into her problems, and motivation (check item) as being greater. They likewise judged her social adjustment as more adequate, more often recommended psychotherapy or psychoanalysis without adjunctive forms of treatment (drugs, shock), and indicated that they would strive to accomplish a more extensive change in the patient's character structure. Presenting Mrs. D. with high motivation also accounted for a more favorable prognosis with treatment and a greater willingness on the part of the therapist to take her into treatment. Finally, as hypothesized, therapists claimed a warmer attitude toward the patient when high motivation was communicated.

Effects of Perceived Patient Motivation

The preceding comparisons were made in terms of the level of motivation communicated in the case material which did not necessarily coincide with the patient's motivation as actually perceived by the respondents. For the male case, there was high agreement between the kind of motivation communicated and the kind of motivation perceived by respondents. For the female case, however, the disparity between communicated and perceived motivation was greater, as evidenced by 13 shifts (16%). Ten of these respondents "misread" the case history intended to convey low motivation by estimating the patient's motivation to be relatively high. Consequently, it was thought advisable to determine the relationship between perceived motivation and other variables, and chi square tests were performed between the respondent's estimate of the patient's motivation and each of the other 23 multiple-choice items.

For the male case, questionnaire items already cited as being significantly related to communicated motivation were again significantly related to perceived motivation. In addition, five other items were significantly

related to perceived motivation at or beyond the .05 level. Respondents who perceived Mr. J. to be highly motivated also tended to estimate social adjustment to be more adequate, suggested more frequent therapeutic sessions, stated that they would be less willing to alter the kind of therapy they usually conduct, estimated the prognosis to be more favorable with treatment, and found it easier to empathize with the patient.

For the female case, estimates of greater ego strength, more insight, better social adjustment, more favorable prognosis with treatment, greater willingness to accept the patient in treatment, and warmer therapist attitudes accompanied perceived motivation, as they did communicated motivation. On the other hand, perceived high motivation (unlike communicated high motivation) was not

significantly related to form of treatment or to the extensiveness of attempted change in the patient's character structure. However, respondents who perceived Mrs. D. as highly motivated for therapy also tended to judge her as being more emotionally mature and less seriously disturbed.

Effects of Therapists' Attitudes

In order to test the second hypothesis, the respondents' stated attitudes toward the patient were compared, by chi square, with the 23 remaining multiple-choice items.

For the male case, eight items varied significantly with the therapist's attitude toward the patient, and only one of these (willingness of the therapist to alter his usual way of conducting therapy) was in the opposite of the predicted direction. Briefly, a positive thera-

TABLE 2
SUMMARY OF STATISTICAL COMPARISONS BETWEEN PATIENTS' MOTIVATION FOR THERAPY,
THERAPISTS' ATTITUDE, AND OTHER QUESTIONNAIRE ITEMS
(*N* = 82)

Therapists' Estimates	Male Case		Female Case	
	Motivation ^b	Therapist Attitude	Motivation ^b	Therapist Attitude
Diagnosis	<i>NS</i> ^a	.02	<i>NS</i>	<i>NS</i>
Ego strength	.001	.01	.05	<i>NS</i>
Anxiety	.05	<i>NS</i>	<i>NS</i>	<i>NS</i>
Insight	.02	.001	.01	.05
Emotional maturity	<i>NS</i>	<i>NS</i>	.01	<i>NS</i>
Social adjustment	.05	<i>NS</i>	.01	.05
Degree of disturbance	<i>NS</i>	<i>NS</i>	.01	<i>NS</i>
Similarity to patients in therapy	<i>NS</i>	<i>NS</i>	<i>NS</i>	<i>NS</i>
Kind of treatment	<i>NS</i>	<i>NS</i>	<i>NS</i>	<i>NS</i>
Motivation		<i>NS</i>		.001
Chances of acting out	.1	<i>NS</i>	<i>NS</i>	<i>NS</i>
Frequency of sessions	.05	<i>NS</i>	<i>NS</i>	<i>NS</i>
Extensiveness of change in character structure	<i>NS</i>	<i>NS</i>	.1	<i>NS</i>
Length of treatment	<i>NS</i>	<i>NS</i>	<i>NS</i>	<i>NS</i>
Permissiveness	<i>NS</i>	<i>NS</i>	<i>NS</i>	<i>NS</i>
Encourage free association	<i>NS</i>	<i>NS</i>	<i>NS</i>	<i>NS</i>
Recommendations to patient	<i>NS</i>	<i>NS</i>	<i>NS</i>	<i>NS</i>
Change of usual therapy	.01	.05	<i>NS</i>	<i>NS</i>
Prognosis without therapy	<i>NS</i>	.05	<i>NS</i>	<i>NS</i>
Prognosis with therapy	.05	.02	.001	.05
Willingness to accept patient in Rx.	.01	.001	.001	.001
Ease of empathizing with patient	.01	.001	<i>NS</i>	.001
Environmental stress	<i>NS</i>	<i>NS</i>	.1	<i>NS</i>
Therapist's attitude	<i>NS</i>		.001	

^a *NS*—not significant at .10 level (chi square, two-tailed tests).

^b Patient's motivation as perceived by therapist.

pist attitude was found to be associated with: a diagnosis of *psychoneurosis* rather than one of psychosis or personality disorder, estimates of greater ego strength and insight, less willingness to alter his usual therapeutic procedure, more favorable prognostic estimates with or without treatment, greater willingness to accept the patient for treatment, and greater ease of empathizing with the patient.

For the female case, six items showed a significant relationship with attitude at the .05 level, and all differences were in the predicted direction. Therapists who claimed a more positive attitude toward Mrs. D. considered her to possess greater insight into her

problems, more adequate social adjustment, and higher motivation for therapy. Warmer attitudes toward this patient were also accompanied by estimates of a more favorable prognosis with treatment and greater ease of empathizing with the patient.

In addition to the statistically significant results already reported, more than 75% of the nonsignificant comparisons showed trends in the predicted direction. The figures for these comparisons are given in Table 2.

Additional Analyses

The theoretical possibility that basic dimensions might underlie the observed relationships among questionnaire items led to

TABLE 3
INTERCORRELATION AMONG SELECTED ITEMS
(*N* = 75 to 82)

Item	Ego Strength	Insight	Social Adjustment	Motivation	Alter Rx.	Accept Patient	Prognosis	Empathy
Male Case								
Insight	.30							
Social adjustment	.61	.28						
Motivation	.32	.50	.27					
Change of usual therapy	-.30	-.31	-.06	-.34				
Prognosis with therapy	.46	.32	.32	.26	-.51			
Willingness to ac- cept patient	.34	.30	.20	.40	-.30	.40		
Ease of empa- thizing with patient	.43	.14	.25	.22	-.02	.22	.28	
Therapist's attitude	.52	.23	.10	.22	-.29	.42	.55	.44
Female Case								
Insight	.49							
Social adjustment	.37	.35						
Motivation	.40	.65	.37					
Change of usual therapy	-.27	-.15	-.07	-.18				
Prognosis with therapy	.41	.37	.28	.55	-.52			
Willingness to ac- cept patient	.19	.39	-.08	.46	-.21	.68		
Ease of empa- thizing with patient	.08	.35	.20	.25	-.01	.32	.40	
Therapist's attitude	.30	.40	.31	.59	-.07	.48	.54	.44

an attempt to isolate such factors. Since a factor analysis employing the intercorrelation of all 27 items seemed impractical, only the 9 items already found to yield significant comparisons were utilized. For each case, these 9 items were intercorrelated (product-moment coefficients of correlation). A centroid factor analysis was performed for each patient separately, three factors were extracted in each analysis, and both factor analyses were then rotated to an oblique simple structure.

The similarity of patterns of the intercorrelation matrices (see Table 3) for each case suggests that certain items are consistently related to other items and that these relationships are rather stable. That is, patient differences as represented by the two cases in this investigation have little effect upon the degree to which various therapist estimates and judgments are correlated, and the observed association between items can be assumed to be mainly a function of therapist variables.

The three factors extracted for each patient are difficult to interpret, and the findings for the two patients often failed to parallel each other. However, common to both factor analyses was the emergence of what might be termed a "good patient" factor. For Mr. J. this factor includes high loadings on items which give the following description: the therapist is willing to accept a well motivated, insightful patient for his usual treatment procedure. For Mrs. D. high loadings occur on such items as: the therapist perceives the patient as having insight, high motivation for therapy, and is easy to empathize with; the therapist has a warm attitude toward the patient and is eager to accept her for treatment.

DISCUSSION

Since this investigation is not based on observations of psychotherapists in their natural setting, questions may be raised about the correlation between their behavior responding to questionnaires and their behavior in actual psychotherapy when they have the opportunity for direct acquaintance with the patient. As one of the authors (Strupp, 1958) pointed out in a somewhat different context,

It is not contended that the behavior of the psychiatrists in an initial interview with a "real" patient bears a one-to-one relationship to his performance in the experiment. . . . What is asserted is that the evaluative processes, as they were studied under experimental conditions, are sufficiently similar to the "normal" behavior of the therapist to generate important hypotheses about his behavior in a first interview. Hypotheses thus derived are valuable if they raise further questions for research, and if they can be applied to practice and training in psychotherapy (p. 64).

Asch (1952) comes to a similar conclusion in evaluating the generalizability of his experimental findings which were based on descriptions rather than on observation of actual persons: "certain processes occurring in response to indirect data about persons not present are also at work in reaction to actual persons" (p. 219). Nevertheless, the possibility that concrete relations with persons introduce additional variables cannot be dismissed lightly.

Regarding our findings, it is seen that certain therapist estimates pertaining to perceptions of, and treatment plans for, the patient are related to the therapist's professed attitude toward the patient. Estimates of diagnosis, ego strength, inclination to modify therapy, prognosis without treatment, social adjustment, and motivation for therapy appear to vary with the therapist's attitude and the patient's degree of disturbance (it will be recalled that the female patient was less seriously disturbed than the male patient).

It should be pointed out that only 10 of the 24 multiple-choice items yielded results which were consistently statistically significant. At the same time, the fact that these results were obtained for two patients adds weight to the evidence. Furthermore, the fact that the vast majority of statistically non-significant comparisons were in the predicted direction, coupled with the respondents' frequently expressed claim that the written case histories did not allow them to use all the cues they ordinarily employ, suggests that many of the relationships are probably underestimated.

The direction of causation obviously remains indeterminate. Thus, it is not clear whether the therapist's attitude "causes" the character of his clinical estimates, whether

the therapist feels more favorably disposed toward patients who are better therapeutic risks to start with, or whether both possibilities are more complex functions of other variables. It seems clear, however, that insight, or the capacity for insight, is a key variable entering into prognostic estimates, the therapist's eagerness in accepting the patient for treatment, and his (self-rated) ability to empathize with the patient. Therapists tend to profess a more positive attitude towards a patient who meets these criteria. On the other hand, the data suggest that the patient's motivation for therapy influences the therapist's attitude in the predicted direction only when the degree of disturbance is not too great. In other words, a therapist "likes" a patient who has the capacity for insight and is eager to be helped provided the clinical evidence supplies no contraindications.

These findings show some interesting parallels to the ingenious experiments of Asch (1946) on the formation of impressions of personality. Among many other findings, Asch reports that "a central trait determines the content and the functional place of peripheral traits within the entire impression" but that central characteristics are "themselves affected by the surrounding characteristics" (p. 276). In his studies, the "warm-cold" dimension appeared to occupy such a central position. In this investigation, motivation for therapy similarly seems to determine a number of clinical estimates as well as the therapist's attitude. (For further evidence on this point, see the study by Strupp & Williams, 1960). Our results also seem to corroborate Asch's conclusions that one can see another person only as a unit, that qualities enter into a particular relation and become organized, and that the system of the traits points to a necessary center. Altogether it seems that application of Asch's technique to the investigation of the formation of clinical judgments presents a fascinating research task.

Regardless of patient's motivation for therapy, however, therapists profess a more positive attitude towards patients who are less severely disturbed. While this investigation has nothing to say about therapeutic out-

comes, it dovetails with the findings of a number of such studies summarized by Luborsky (1959): "Those who stay in treatment improve; those who improve are better off to begin with than those who do not; and one can predict response to treatment by how well they are to begin with" (p. 324). We may add, on the basis of our findings, that therapists are more enthusiastic about treating patients whom they view as having a better prognosis—perhaps, because they are perceived as having greater capacity for profiting from psychotherapeutic procedures which rely both on the patient's intellectual as well as emotional understanding. Our first hypothesis may thus be recast: A "warm" therapist attitude toward a patient is more likely to be present when the patient is well motivated for therapy, provided the severity of his emotional disturbance is not too great.

The one result which ran counter to the prediction showed a negative relationship between a favorable therapist attitude and his willingness to alter his usual method of therapy. In the context of other findings here reported this disinclination to alter his usual method of therapy seems to reflect the therapist's conviction that the patient can profit from the kind of psychotherapy the therapist employs most often and has the greatest faith in and that there is no need to press into service other techniques and therapeutic adjuncts—the use of which may, at times, signify a subtle rejection of the patient as a person.

Estimates of the patient's insight into his problems showed the highest degree of association with the patient's motivation for therapy and the therapist's attitude. This finding supports the notion that estimates of insightfulness are highly descriptive of the good (potential) learner as far as psychotherapeutic achievement is concerned. It also substantiates the belief that psychotherapeutic success requires, among other things, the patient's ability to achieve insight and the therapist's ability to foster a warm, understanding, and accepting environment. Fiedler (1950) found these factors to be mentioned prominently in therapists' descriptions of the ideal therapeutic relationship, and he observed that agreement on these variables was

higher among experienced therapists regardless of theoretical orientation than among experienced and inexperienced therapists of like orientation.

The demonstrated relationships between therapists' attitudes and their clinical evaluations and treatment plans in general support the second hypothesis. This may also be interpreted as discrediting the conception of an impersonal, mirror-like therapist, a view which has already become outmoded. By drawing attention to these associations, this study supports the increasing emphasis given to the effects of personality variables in the therapist upon his clinical operations. The extent to which these variables play a part in the therapist's intricate work remains to be explored much more exhaustively, but it is clear that they must come to occupy a more central place in selection, training, and therapeutic practice. From a methodological point of view, this study demonstrates the feasibility and potential value of quasiclinical experimental designs, which permit the introduction of controls not otherwise possible.

SUMMARY

This investigation deals with an exploration of two hypotheses positing an association between therapists' attitudes toward patients and clinical evaluations and treatment plans as a function of the patient's motivation for therapy. Eighty-two respondents, all of whom were practicing medical psychotherapists, read two case histories of neurotic patients. Alternate forms of the case histories were employed to vary systematically the patient's motivation for therapy. Respondents recorded their clinical impressions, estimates, and attitudes by means of a 27-item questionnaire which followed each case presentation.

The hypothesized association between therapists' attitudes toward the patient and their perceptions of, and treatment plans for, the patient appear in general to be substantiated. It is pointed out that the direction of causation remains indeterminate and that the degree of relationship between the therapist's

"usual" performance and his performance in the present experimental situation is unknown. Nevertheless the evidence is believed to support the assertion that personality factors of the therapist are an integral part of his clinical judgments and therapeutic procedures. A better understanding of these relationships might make an important contribution to more effective clinical practice as well as aid in the selection of candidates for training in psychotherapy.

REFERENCES

- ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, 41, 258-290.
- ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
- BUTLER, J. M. The interaction of client and therapist. *J. abnorm. soc. Psychol.*, 1952, 47, 366-378.
- FIEDLER, F. E. A comparison of therapeutic relationships in psychoanalytic, nondirective and Adlerian therapy. *J. consult. Psychol.*, 1950, 14, 436-445.
- GITELSON, M. The emotional position of the analyst in the psychoanalytic situation. *Int. J. Psycho-Anal.*, 1952, 33, 1-10.
- JACKSON, D. D. Countertransference and psychotherapy. In Frieda Fromm-Reichmann & J. L. Moreno (Eds.), *Progress in psychotherapy—1956*. New York & London: Grune & Stratton, 1956. Pp. 234-238.
- LUBORSKY, L. Psychotherapy. In P. R. Farnsworth & Q. McNemar (Eds.), *Annu. Rev. Psychol.* Vol. 10. Palo Alto, Calif.: Annual Reviews, 1959. Pp. 317-344.
- MENSH, I. N. Research in counseling and psychotherapeutic process. In D. Brower & L. E. Abt (Eds.), *Progress in clinical psychology*. Vol. 2. New York & London: Grune & Stratton, 1956. Pp. 340-360.
- PARLOFF, M. B. Some factors affecting the quality of therapeutic relationships. *J. abnorm. soc. Psychol.*, 1956, 52, 5-10.
- STRUPP, H. H. The psychotherapist's contribution to the treatment process. *Behav. Sci.*, 1958, 3, 34-67.
- STRUPP, H. H., & WILLIAMS, JOAN V. Some determinants of clinical evaluations of different psychiatrists. *AMA Arch. gen. Psychiat.*, 1960, 2, 434-440.
- WINDER, C. L. Psychotherapy. In P. R. Farnsworth & Q. McNemar (Eds.), *Annu. Rev. Psychol.* Vol. 8. Palo Alto, Calif.: Annual Reviews, 1957. Pp. 309-330.
- WOLBERG, L. R. *The technique of psychotherapy*. New York: Grune & Stratton, 1954.

(Received August 3, 1959)

OPERANT CONDITIONING OF HOSTILE VERBS IN RELATION TO EXPERIMENTER AND SUBJECT CHARACTERISTICS¹

DONALD C. FERGUSON AND ARNOLD H. BUSS

University of Pittsburgh

Virtually all operant verbal conditioning studies have used neutral materials. One of the exceptions is an experiment by Buss and Durkee (1958), who used a modification of Taffel's procedure (1955). A list of past tense verbs was compiled, and judges sorted them into three categories: neutral, mildly hostile, and intensely hostile. Three verbs, one from each category, were placed on a card, and Ss were instructed to make up a sentence using one of the verbs. There were 60 such cards (trials). One group of Ss was reinforced for using intensely hostile verbs, and another group was reinforced for using neutral words. It was found that intensely hostile words were learned at a faster rate than neutral words. There were also sex differences in frequency of using intensely hostile words.

This study was followed up by Binder, McConnell, and Sjöholm (1957), who used only the neutral and mildly hostile verbs. There were two groups of Ss, both reinforced for using hostile verbs. One group was run by a large, aggressive man and the other by a diminutive, mild, nonassertive woman. The group run by the "neutral" woman learned significantly better than the group run by the aggressive man.

Evidently characteristics of the *E* affect learning of hostile material. However, it is not clear from the Binder et al. (1957) study which characteristics are important. One *E* was large, male, and aggressive; the other *E* was small, female, and neutral. It is doubtful that size of *E* is a major variable, which leaves aggressiveness and sex. Which of these variables retarded learning, aggressiveness or

masculinity? Perhaps both affected learning. The Binder et al. (1957) experiment does not furnish an answer because aggressiveness and gender of the *E* were confounded. The present study investigated these variables separately, i.e., a male and a female *E* each played an aggressive or a neutral role with Ss. In addition, sex of Ss was a variable in an attempt to discover interactions between characteristics of the *E* and gender of the *S* in the operant verbal conditioning of hostile material.

METHOD

Experimental Design

There were four groups of Ss. Two groups were run by a man and two groups by a woman. The male *E* was the senior author, and the female *E* was his wife.² Each *E* played a neutral role with one group and an aggressive role with the other. Pre-experimental rehearsal helped to make enactment of each role similar for the two *Es*.

The neutral *E* was patient, calm, and courteous in an attempt to establish a nonhostile experimental climate. The aggressive *E* was brusque and unfriendly. The *S* was made to wait if early or on time and criticized if late. The *E* was impatient and tended to scowl and sneer. Reports from Ss at the termination of the experiment revealed that the two roles (neutral or aggressive) were enacted appropriately.

Procedure

The procedure was the same as that of Binder et al. (1957). The *S* was presented with a sample card on which was typed the verbs WALKED and RAN, and the pronouns HE, THEY, SHE. He was instructed to make up a sentence using one of the verbs and one of the pronouns. Then he was presented with a series of 140 similar cards, each card constituting a trial. On these cards one verb was neutral and one was mildly hostile.

² The writers wish to thank Elsie Ferguson for serving as an *E*.

¹ This report is based on a master's thesis by the senior author under the direction of the junior author.

The first 20 trials were used to establish an operant level of responding. Starting with Trial 21 the *E* said "Good" each time a hostile verb was used. This procedure was the same for all *Ss*.

Subjects

The *Ss* were 80 college students recruited from psychology classes. The *Ss* were randomly assigned to the four experimental groups, with the restriction that half the *Ss* in each group were men and half were women.

RESULTS

An examination of the learning curves revealed no systematic sex differences. Therefore the learning curves of male and female *Ss* were combined. The data bearing on aggressiveness of the *E* are presented in Fig. 1. The neutral *E* curve is a composite of the data from the male and female *Es*, and the aggressive *E* curve combines groups run by the male and the female *Es*. The aggressive *E* curve is smoother and has a steeper slope than the neutral *E* curve. The superiority of learning with a neutral *E* is especially evident during the last 60 trials (Blocks 5, 6, and 7).

The data bearing on gender of the *E*, presented in Fig. 2, indicate that the sex of the *E* has little effect on learning. The female *E* curve starts at a slightly higher level than the male *E* curve. However, this difference disappears during the second block of trials, and thereafter the curves are similar.

None of the learning curves presented in the two figures is smooth, but the irregularities do not fall into any pattern. In this respect they resemble the curves obtained in most operant verbal conditioning studies, in

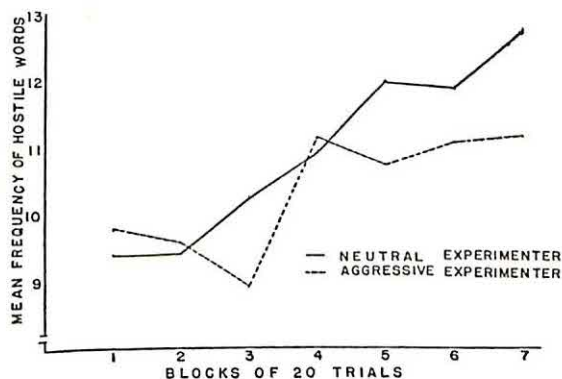


FIG. 1. Conditioning of hostile verbs with a neutral or aggressive *E*.

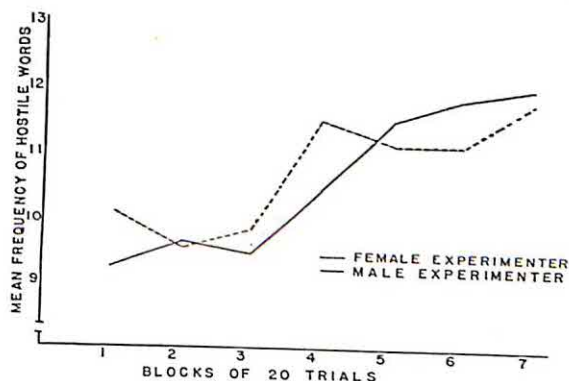


FIG. 2. Conditioning of hostile verbs with a male or female *E*.

which uncontrolled variables appear to be affecting the conditioning process. In this context neither the shape of the curve (which reflects random irregularities) nor the overall height of the curve is of particular importance. Overall height is determined not only by increments in learning but also by initial level of responding. Therefore, the crucial measure of conditioning is the increment in learning, i.e., the net increase in frequency of hostile verbs above the operant level of responding. The score used was: frequency for last 20 trials minus frequency for the first (operant) 20 trials.³

An analysis of variance of these scores is presented in Table 1. The only significant *F* is for aggressiveness of the *E*. Sex of the *E*, sex of the *S*, and interactions among the variables were not significant. This analysis of

³ One of the writers (Buss) has conducted several experiments on operant verbal conditioning with both neutral and hostile materials. These studies have yielded learning curves that vary not only from one *S* to the next but from one sample to the next. The lack of smoothness in curves has not obliterated the presence of group differences, probably because such differences were large, but it warns against placing much faith in intertrial variations. An analysis of trend, such as that used by Binder et al. (1957), takes advantage of variations among blocks of trials. Since such intertrial variations tend to be unreliable, they probably consist mainly of error variance. Therefore, trend analysis was not used in testing the significance of the present data. Rather, an analysis of variance of the net increase in response frequency was employed. The net increase measure seems best because it is the most conservative and the most stringent measure of learning, since it does not take advantage of random variations within blocks of trials.

TABLE 1

ANALYSIS OF VARIANCE OF DIFFERENCE SCORES FOR
FREQUENCY OF HOSTILE VERBS

Source	df	Mean Square	F
Aggressiveness of <i>E</i>	1	66.61	4.24*
Sex of <i>E</i>	1	30.01	1.91
Sex of <i>S</i>	1	1.01	.06
Sex of <i>E</i> × Sex of <i>S</i>	1	10.51	.67
Sex of <i>E</i> × Aggressiveness of <i>E</i>	1	2.11	.13
Sex of <i>S</i> × Aggressiveness of <i>E</i>	1	2.11	.13
Sex of <i>S</i> × Sex of <i>E</i>			
× Aggressiveness of <i>E</i>	1	.32	.02
Within groups	72	15.70	
Total	79		

* $p < .05$.

variance confirms the trends noted in the learning curves. Only aggressiveness of the *E* significantly affects learning, an aggressive *E* tending to retard conditioning.

Since the curves in Figs. 1 and 2 differ in overall height, another score was used: total frequency of hostile verbs. An analysis of variance of these scores yielded no significant *F*. Evidently none of the variables influenced total frequency of hostile verbs.

DISCUSSION

Binder et al. (1957) found that an aggressive male *E* retarded learning in comparison to a neutral female *E*. The results of the present study indicate that the effective variable is not the sex but the aggressiveness of the *E*. Operant conditioning of hostile verbs is slower with an aggressive *E* than with a neutral *E*. In attempting to explain this finding two possibilities should be considered. First, when the *E* was aggressive, he may have been a less potent reinforcing agent. His saying "Good" would be less reinforcing than the neutral *E*'s "Good" and would lead to slower learning.

An alternative explanation is that *E*'s being aggressive angered the *S*. The *S* could not aggress openly against the *E*, because of his higher status as an *E*. Therefore, it would be necessary to repress or suppress aggressive tendencies, and such inhibition would generalize to the hostile verbs. The *S* would then tend to inhibit verbalizing the hostile choice,

and this would retard conditioning. Which of these two explanations is correct will be decided by studies in which other kinds of material are conditioned and the *E* assumes nonneutral roles other than aggressiveness.

Two of the groups in the present study may be compared with their counterparts in the Binder et al. (1957) study: the group run by the neutral female *E* and the group run by aggressive male *E*. The neutral female *E* curves in the two experiments are similar, with the curves of the present study starting slightly higher and ending slightly higher; the slopes appear almost identical. The aggressive male *E* curves start from the same operant level of responding, but the curve of the present study has a steeper slope, ending at a higher point. Considering the variability in these curves from one block of trials to the next, it is doubtful that these differences in the aggressive *E* curves are significant.

In both studies the slope of the neutral *E* curve was significantly steeper than that of the aggressive *E* curve, i.e., the increment in frequency of hostile verbs was significantly greater for the *S*s run by a neutral *E*. Concerning the total frequency of hostile verbs, the two experiments differ. Binder et al. (1957) found a significant difference between the neutral *E* and aggressive *E* groups in total frequency of hostile verbs, whereas the present investigators did not, although the difference was in the same direction. However, as noted earlier total frequency is influenced not only by increments in learning but also by initial level of responding and by irregularities in the learning curves. Examination of Fig. 1 reveals that the curve of the aggressive *E* started higher and because of irregularities in the curve, it overlapped the neutral *E* curve at the fourth block of trials. These two facts probably account for the difference in total frequency of hostile verbs not being significant in the present study. However, since the crucial dependent variable is *increment* in frequency of response, not height of the curves, this discrepancy between the Binder et al. (1957) study and the present one seems unimportant. Such minor differences must be expected so long as we are dealing with curves that are subject to random irregularities. With respect to the major

finding—that an aggressive *E* retards conditioning of hostile verbs—the two studies are in agreement.

SUMMARY

This experiment investigated operant conditioning of hostile verbs in relation to aggressiveness of the *E*, sex of the *E*, and sex of the *S*. The *Ss* were presented with two verbs and three pronouns on a card. One verb was hostile and the other neutral. The *Ss* made up a sentence using one verb and one pronoun. They were reinforced for using hostile verbs. It was found that only the aggressiveness of the *E* significantly affected

conditioning: an aggressive *E* retards learning in comparison to a neutral *E*.

REFERENCES

- BINDER, A., McCONNELL, D., & SJOHOLM, NANCY A. Verbal conditioning as a function of experimenter characteristics. *J. abnorm. soc. Psychol.*, 1957, **55**, 309-314.
- BUSS, A. H., & DURKEE, ANN. Verbal conditioning of hostile and neutral verbs. *J. consult. Psychol.*, 1958, **22**, 415-418.
- TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496-501.

(Received August 9, 1959)

THE EFFECT OF A STRUCTURED INQUIRY ON RORSCHACH SCORES¹

MELVIN ZAX AND GEORGE STRICKER
University of Rochester

In spite of the fact that the inquiry is an integral part of the Rorschach examination there is little unanimity among experts concerning the proper way to conduct it, nor is much empirical data available on the effects of various procedures. All prominent Rorschach theorists agree that the examiner should adopt a nonleading approach when conducting the inquiry. When they operationalize their recommendations, however, it is apparent that they disagree on what constitutes "leading the subject."

Klopfer and Kelley (1946) advocated procedures such as offering alternative concepts or comparing a previously given percept to a later one in order to establish the presence or absence of a specific determinant. In probing for movement they also sanctioned questioning about the positions of the figures seen. As a last resort in ruling on the use of color they agreed with Beck (1950) in employing such questions as, "Suppose this were exactly the same shape as now, but gray or black, would you still think it is . . . ?"

Klopfer, Ainsworth, Klopfer, and Holt (1954) cautioned against suggesting any particular determinant to subject (*S*) and, as a way of avoiding this, advised sticking closely to the terms and language employed by *S* himself. They were also critical of the practice of comparing the Rorschach plate with the small achromatic designs on location charts which was recommended by Piotrowski (1957). Klopfer et al. (1954), however, did advocate questions about the position of the percept and how it was seen when probing for movement.

Among the most conservative of the Rors-

chach testers with regard to the inquiry are Rapaport, Gill, and Schafer (1946) who recommended conducting the inquiry after each card and with the blot not present. Questioning is kept to an absolute minimum. Thus, only what is spontaneously perceived is scored, and the record is kept free of rationalization and secondary elaboration. Sarason (1954) likewise favored a minimum of questioning, and apart from a reading back of the response with the opening query, "What was it in the card that reminded you of that?" only the clarification of any ambiguity in the client's verbalizations was allowed.

Data on the differential effects of such procedures is sparse indeed. Gibby and Stotsky (1953) have provided some information on the effects of the inquiry on Rorschach scores by comparing the scoring of the free association on patients' records to the scoring of the free association plus the inquiry. The type of inquiry used was not standardized. It was found that the indices could be differentially ordered for amount of change with *F* ranking first and *FV*, *FC*, *CF*, *M*, *YF*, *V*, and *Y* following in that order. Of course, *F* decreased while all the other indices increased, and the changes for all but *V* and *Y* were statistically significant.

Baughman (1958) proposed a new method of conducting the inquiry whereby the standard inkblots are compared systematically to a set of especially prepared blots in which color, shading, figure-ground contrast, and complexity of form are varied. *Ss'* responses to a standard series of questions reveal which of the determinants have been most influential. Contrasting the scoring of records elicited by this procedure with records in which the traditional inquiry was done, Baughman

¹The authors are grateful to Dorothea Brayer, Florence Budin, A. Cooper, J. H. Weiss, and D. L. Wolitzky who served as examiners in this study.

found that significantly more shading determinants were given in response to the paired comparison technique. However, there were no significant differences in the number of scorable color responses elicited by either technique.

In a study closely related to the one reported herein, Klingensmith (1956) investigated the effects of the amount of structure used in the inquiry on the quantity and type of indices reported. The scoring of the free association alone represented the least structured procedure. Next the scoring of the free association plus the traditional inquiry was computed. Thirdly, Ss were asked to return to the cards and give whatever determinants besides shape were influential in their responses and a separate scoring was done for this procedure. Finally, the determinants were described to S and he was told to report those he used. Again the records were scored using information elicited by this technique. When the scoring from the various procedures was compared, it was found that *T*, achromatic color, and *V* showed the greatest increase with the increase in structure while *M*, *FM*, *C*, and *Fm* showed less tendency to increase. The increase in *M* was insignificant at all levels. Unfortunately Klingensmith limited his Ss to two responses per card and only analyzed for 15 responses from each of the records taken.

This study is essentially a replication of that of Klingensmith with a slight procedural modification and with the entire record of each S being considered. This is looked upon as an exploratory study of the extent and type of change in scoring which results from a frank description of determinants concerning which the inquiry is being done. It should thereby provide data on the reliability of the indices as a function of the procedure by which they were elicited.

METHOD

Ss for this experiment were secured by five examiners who each administered 15 tests as part of the requirement for credit in a graduate course in Rorschach administration. Before any examiner was permitted to begin actual administration of the test, he must have participated in all of the classroom instruction and have been observed and supervised in at least two practice examinations. Examiners

were instructed to use only Ss who were between 16 and 50 years of age and who had never been hospitalized for an emotional disorder. In the sample which was collected the age ranged from 16 to 47 years with the mean age being 25.79 years. The mean education level of these Ss was 14.07 years (about two years of college) and the majority (55%) were students when tested. The sample comprised 30 males and 45 females.

Procedurally, the test administration was introduced by the following statement:

I have a set of ten inkblots that I'd like you to look at. They are made simply by dropping ink on a sheet of paper and folding the paper so that the ink spreads out in a chance way. Different people can see different things in these blots and I'm interested in what you can see in them. So, tell me whatever you see as soon as you see it. (After the first card has been returned add: From now on when you can't see any more put the card down as a signal that you've finished.)

The free association period was followed by the inquiry which was conducted in the manner described by Sarason (1954), since he gives the most precise directions and also because his approach represents one of the least structured in current use. Subsequent to this inquiry, a second inquiry was done which was prefaced by this statement:

Inquiry II. The first time we went through the blots you may not have been certain just what I was getting at. The point is there are a number of features of the inkblots which may possibly help to determine the many things one can see in them. The shape or contour of the blot is obviously one of these, but other possibilities are the color of the blot, the shading within it, the suggestion of a three dimensional effect or perspective, the appearance of a textural quality to the surface and the illusion that the figures seen are in motion or assuming some position. To ensure that none of these possibilities have been missed in going over your responses, I'd like to take one more look at them and ask you to make certain that I understand in each case exactly what it was that caused the blot to look as it did to you when you first saw it. Let's go over the cards once more and you tell me what about the blot led you to see it the way you did, when you first looked at it.

In this second inquiry Ss' original responses and their remarks from the first inquiry were read back to them and they were invited, but not encouraged, to give any additional determinants which applied.

For each record a separate scoring was done for the free association, the free association plus the first inquiry, and the free association plus both inquiries. Approximately half of the records were scored by each of the authors, both of whom have had extensive experience with the test. Fifteen records, three being taken from each of the examiners, were scored in common to provide a reliability

check. Scoring of determinants was done according to the method prescribed by Klopfer et al. (1954) except that Beck's indices for shading were adopted.

RESULTS

To insure the reliability of the scoring system between the two examiners a series of 27 rhos was computed. These compared the scoring of the authors on nine indices or combinations of indices elicited, for each of the three aspects of the testing situation (free association and the two inquiries). The indices studied were *M*, *FM*, *Fm*, *FC*, *CF* + *C*, *FY*, *YF* + *Y*, *FV*, and *FT* + *TF* + *T*. Seventeen rhos were greater than .90, eight fell between .80 and .90, and only two fell below .80, one being .74 and the lowest being .68 (for the scoring of *FC* on the free association in which case there were relatively few scorings made and small differences had a great effect on the overall rho). On the basis of these figures it was felt that the scoring system was a reliable one.

Since the experiment was concerned with the effect of inquiry structure on the elicitation of determinants, each *S* was used as his own control and the scorings of his free association, free association plus inquiry, and free association plus both inquiries were compared. To accomplish this, the total number of each index given by *S* in response to the entire test procedure was computed. The percentage of this total which was scorable at each of the three levels of structure was then calculated, and statistical comparisons were

made between these figures. In this case the null hypothesis was that the levels of inquiry structure (the free association period was considered to represent the complete absence of such structure) had no effect on the tendency to produce the various determinants. By chance it would be expected that *Ss* would either give about as many indices of any type on one phase of the test as on others or that those who give the most at a single level of structure are counterbalanced by others who are most responsive at a different level. The statistical test used was the Friedman non-parametric analysis of variance for repeated measurements of the same *S* as described by Siegel (1956).

A series of eight such analyses was done for the indices *M*, *FM*, *Fm*, *FY*, *YF* + *Y*, *FV*, *FC*, and *CF* + *C*. No analysis was done for *FT* + *TF* + *T* since so few texture responses were given. In the case of each determinant, of course, analyses were restricted to those *Ss* giving them so that the *N* for each index varies somewhat. Table 1 presents the mean percentages of the indices given at each level of structure, the mean number of responses involving these indices at each level, and the χ^2 for each analysis which is referable to the standard chi square table (in this case at two degrees of freedom).

The analyses in Table 1 indicate that for all of the indices save *FV* there are highly significant differences between the percentages elicited at the various levels of structure. The findings with regard to the movement

TABLE 1
MEAN PERCENTAGES AND NUMBER OF RESPONSES AT EACH LEVEL OF STRUCTURE
AND RESULTS OF NONPARAMETRIC ANALYSES OF VARIANCE

Index	<i>N</i>	Mean % Free Association	Mean No. of Responses	Mean % Inquiry I	Mean No. of Responses	Mean % Inquiry II	Mean No. of Responses	χ^2 *
<i>M</i>	72	83	3.35	16	0.65	1	0.05	83.46
<i>FM</i>	73	63	3.65	32	1.88	5	0.28	62.82
<i>Fm</i>	41	50	1.03	48	0.99	2	0.05	24.74
<i>FY</i>	75	4	0.31	58	3.81	38	2.46	63.55
<i>YF</i> + <i>Y</i>	41	13	0.27	60	1.30	27	0.57	15.01
<i>FV</i>	33	29	0.62	48	1.05	23	0.51	2.80
<i>FC</i>	67	12	0.38	71	2.28	17	0.56	51.77
<i>CF</i> + <i>C</i>	60	28	0.91	62	2.05	10	0.32	39.11

* For two degrees of freedom χ^2 of 9.21 is significant at .01.

TABLE 2

TOTAL NUMBER OF Ss RESPONDING WITH THE VARIOUS DETERMINANTS AT EACH LEVEL OF STRUCTURE

Index	N	Free Association	Inquiry I	Inquiry II
<i>M</i>	72	68	27	3
<i>FM</i>	73	67	48	10
<i>Fm</i>	41	28	28	2
<i>FY</i>	75	15	67	57
<i>YF+Y</i>	41	9	31	17
<i>FV</i>	33	13	20	11
<i>FC</i>	67	21	60	23
<i>CF+C</i>	60	33	51	11

categories and the shading determinants, particularly *FY*, are most striking. In the case of *M* only three Ss of the 72 who gave that determinant did so in response to the most structured level of inquiry, and in each case only a single *M* was added. On the other hand 60 of the 75 Ss ultimately giving *FY* failed to give any responses which could be scored as such on the free association, while 57 of the 75 Ss added this determinant in response to the instructions indicating which indices could be given. A complete breakdown of the number of Ss giving each index at each level of structure is found in Table 2.

DISCUSSION

The results of this study indicate that the various determinants (with the exception of *FV*) are characteristically elicited at certain times in the Rorschach examination and that these times vary with the determinant. Furthermore, they indicate that certain indices are more sensitive than others to the increase in inquiry structure. Ss give an average of 2.46 additional *FY* responses in the most structured inquiry while giving only 0.05 additional *M* responses in the same phase of the test. Rank ordering the indices in terms of this tendency to be elicited in the more structured inquiry, *FY* is first followed by *YF* + *Y*, *FV*, *FC*, *CF* + *C*, *FM*, *Fm*, and *M*. This is in substantial agreement with the findings of Gibby and Stotsky (1953) and Baughman (1959), who concerned himself only with color and shading. This is also in close accord with the results of Klingensmith (1956) with the exception that *FV* was not

found to show the greatest increase with the increase in structure. On the contrary, of all the shade-determined indices *FV* showed the least increase on the second inquiry.

Such findings have clear implications for the clinical use of the test. They would suggest that in the absence of a standardized inquiry the *FY* index is highly unreliable. Interpretations based on the number of such determinants present in a record must be made with considerable caution therefore. On the other hand the various movement determinants seem highly reliable. In spite of the fact that what appear to be high percentages of *YF* + *Y*, *FV*, *FC*, and *CF* + *C* are given on the most highly structured inquiry (see Table 1), the increment in terms of average number of such responses per *S* is never more than 0.57 which is quite small. Such differences are perhaps less than one would have been led to expect in the absence of empirical data.

While the findings in this study suggest that most scoring indices are reasonably reliable for clinical use, the implication that a standardized inquiry procedure is unnecessary is not intended. The instability of *FY* alone under varying conditions would argue for a more universally accepted way of conducting the inquiry. Perhaps one of the obstacles to the development of such a procedure is the fact that textbook writers have failed to operationalize the procedure enough. The fear that gross distortions might result from a frank confrontation of *S* with the possibilities among the various determinants may well have been a deterrent to the development of a standard set of instructions. As a result a variety of subtleties are recommended which allow for considerable personal variation in approach.

Baughman (1958) has recognized the need for greater standardization of the inquiry and has proposed a technique for eliciting color and shading responses. Unfortunately, his approach requires the use of six alternate sets of inkblots and a procedure which appears quite time consuming. The use of the kind of structuring employed in the second inquiry in this study might well accomplish the same purpose as his paired-comparison technique, and a comparison of the effects of the two

techniques would be one fruitful direction for further research.

Another worthwhile investigation would involve a comparison of the effects of examiner differences under conditions where the inquiry is an unstructured one and when, by virtue of the instructions given, it is highly structured. Conceivably, the *S* who is given few cues regarding the object of the inquiry might seize on whatever ones are inadvertently offered by the examiner, while the *S* who is given all of the cues he needs through the instructions may remain less sensitive to examiner idiosyncracies. This possibility can be investigated using the data obtained in the reported study, and the authors are currently doing so. If this hypothesis is substantiated, it would lend further support for the establishment of a standard inquiry based on a frank description of the various possible determinants to be found in the blots.

SUMMARY

The effects of inquiry structure on the determinants elicited were studied by administering the standard Rorschach to 75 *Ss*, conducting a relatively unstructured inquiry and then a second one prefaced by a description of the determinants. The free association, free association plus the first inquiry, and free association plus both inquiries were scored separately and these scorings compared. It was found that the various determinants characteristically tended to be elicited in different phases of the test procedure (i.e., most movement responses were scorable on the free association, while most shading and color re-

sponses were only scorable after an inquiry). It was also found that of the determinants studied only the change in number of *FY* on the highly structured inquiry appeared to take on clinical significance. The desirability of a standardized inquiry was discussed, and the possibility of basing it on a frank description of the determinants which are the object of inquiry was considered.

REFERENCES

- BAUGHMAN, E. E. A new method of Rorschach inquiry. *J. proj. Tech.*, 1958, 22, 381-389.
- BAUGHMAN, E. E. The effect of inquiry method on Rorschach color and shading scores. *J. proj. Tech.*, 1959, 23, 3-7.
- BECK, S. J. *Rorschach's test*. Vol. 1. *Basic processes*. New York: Grune & Stratton, 1950.
- GIBBY, R. G., & STOTSKY, B. A. The relation of Rorschach free association to inquiry. *J. consult. Psychol.*, 1953, 17, 359-364.
- KLINGENSMITH, S. W. A study of the effects of different methods of structuring the Rorschach inquiry on determinant scores. Unpublished doctoral dissertation, Univer. Pittsburgh, 1956.
- KLOPFER, B., AINSWORTH, MARY D., KLOPFER, W. G., & HOLT, R. R. *Developments in the Rorschach technique*. Vol. 1. *Theory and technique*. New York: World Book, 1954.
- KLOPFER, B., & KELLEY, D. M. *The Rorschach technique*. New York: World Book, 1946.
- PIOTROWSKI, Z. A. *Perceptanalysis*. New York: Macmillan, 1957.
- RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. 2. Chicago: Year Book Publishers, 1946.
- SARASON, S. B. *The clinical interaction*. New York: Harper, 1954.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

(Received August 19, 1959)

CLINICAL DIAGNOSIS BY THE IPAT MUSIC PREFERENCE TEST

RAYMOND B. CATTELL AND ROBERT E. McMICHAEL

University of Illinois

The present investigation was undertaken with the limited purpose of determining any existing differences in the types of musical preference expressed by a normal control group and various pathologic, psychotic, and neurotic groups. It is an exploratory piece of research in that the measuring instrument used is of quite recent vintage. Though it is convenient, consisting of a single long-playing record which can be responded to on answer sheets by a whole collection of patients, and although the independence of the factor scales has been factor analytically confirmed, very little information has yet come forth regarding its validity or reliability as a clinical instrument.

A search of the literature has shown that even though several studies have employed measures of musical preference not one has directly concerned itself with investigating the determination of diagnostic classification of psychiatric syndrome groups by musical preference. Rather, they have been principally concerned with determining possible variables influencing or associated with musical preference in *normal* subjects (Ss). Among others, such variables as age (Keston & Pinto, 1955; Valentine, 1926), sex (Fay & Middleton, 1941; Kerr, 1943; Keston & Pinto, 1955; Schultz, 1933; Sopchak, 1955), mood (Coroso, 1958; Podolsky, 1954; Sopchak, 1955), "intellectual introversion" (Keston & Pinto, 1955), and music recognition (education) (Keston & Pinto, 1955; Sopchak, 1955), have been investigated. The results of such studies at best suggest some variables that may profitably be controlled when working with measures of musical preference.

In a more dynamic mode, both Cattell (1951, 1957) and Sopchak (1957) have proposed the use of music as a "projective"

device. The former starts with the premise that, "The powerful and immediate connection of musical stimulation with emotional experience, and the many indications that unconscious needs gain satisfaction through this medium, have long pointed to measures of musical preference as effective avenues to deeper aspects of personality" (Cattell & Anderson, 1953b, p. 446). Believing that here, as in his humor studies, psychiatric judgment is not sufficient to justify a priori decisions as to what musical preferences should form the scale content for a particular syndrome, he factored the likes and dislikes for 120 brief musical excerpts and arrived at 11 independent (simple structure) scale dimensions (Cattell & Saunders, 1954). These cannot be explained in terms of historical musical periods or schools but suggest dynamic factors of temperament and emotional need. On this research basis a 100-item, 1-hr. test was recorded and standardized as described elsewhere (Cattell & Anderson, 1953a; Cattell & Anderson, 1953b). Subsequently, a correlation study was carried out between the music factors of the musical preference test and the factors on the 16 Personality Factor Questionnaire (Cattell & Anderson, 1953b). The study showed that although a simple one-to-one relation between the two sets of factors could not yet be established, the correlations are substantial, i.e., the test gets into the same "personality space," and, "the set of 16 PF test factors associated with any one music factor has a psychologically consistent and compatible character . . . in every case" (Cattell & Anderson, 1953b, p. 449).

The same study reports results of music factor score comparisons among psychotics and normals, and it is this part with which the present study continues to further cat-

egories. Such comparisons were suggested by reasoning that if the musical preferences are functions of personality factors then various psychotic and neurotic syndrome groups should show different music factor scores. This reasoning is based on the commonly accepted—however, often untested—proposition that various syndromes are interpretable in terms of extreme patterns of common personality factors (as shown, for example, in Cattell and Scheier's recent studies [1960] of neurotic syndromes with the 16 PF test). Statistically significant results were in fact found, and on the basis of this suggestive, promising, and partially confirmatory evidence of the test's capacity to discriminate among pathologic groups, it was concluded that further investigation, using other samples, was desirable. Consequently, in overall design the present research is similar to that of Cattell and Anderson (1953b). However, because of modifications in the diagnostic categories and sizes of the pathologic groups, in testing procedure and statistical analysis, and in the sex, age, and other characteristics of all the samples, it is by no means a straightforward replication.

METHOD

Design

The IPAT Musical Preference Test, permitting scores on 11 distinct dimensions of choice, was administered to several diagnostically different abnormal groups, in addition to a control or "normal" group. Intergroup comparisons, on their mean scores on each factor, were made between (a) the various pathological groups and the control group, (b) the control group and the total abnormal group, and (c) among the various pathological groups. The design included the additional features that only same sex comparisons were made and that any possible effects attributable to age were first separated out. These comparisons provided the basis for conclusions concerning the discriminating power of the music preference test.

Subjects

The Ss of this investigation consisted of seven groups—organics, sociopathic personalities, schizophrenics, paranoids, affectives, psychoneurotics, and normals. Table 1 presents general, descriptive, identifying data on all the samples used in this investigation. All of the abnormal Ss were patients in Manteno State Hospital, Manteno, Illinois. The criterion used for diagnostic classification was the hospital diagnosis as determined in case conferences. A re-

striction imposed on the abnormal samples was that no severely regressed patients participated in the study. Every patient was able, at least initially, to attend to the music.

Since it seemed undesirable to take students as a control group, as was done in an earlier study, we obtained the cooperation of members of a church social group of roughly similar social status and age range to the mental hospital group. Descriptive data on this sample are, however, limited because these Ss were not required to list any identifying data other than their age and sex. However, educationally, they seemed somewhat more homogeneous, though not necessarily higher, than the abnormals.

Materials

The IPAT Music Preference Test of Personality (Cattell & Anderson, 1953a) was the test instrument used in the present investigation. The test consists of 100 musical excerpts which have been put onto a 12-in., 33 $\frac{1}{3}$ rpm record by RCA Victor. The test was tape recorded in order to facilitate group testing. Good fidelity of reproduction was obtained. A Knight tape recorder and an external speaker system were used for the playing of the music.

Testing Procedure

The normal Ss were tested in one group in a large room at the church. The abnormal Ss were tested in groups as formed by wards. Therefore, various diagnostic groups were represented at a given testing. The taking of the test was strictly a matter of voluntary participation on the part of the patients. There were always from 8 to 10 monitors present who encouraged and supervised the Ss, showing them where to record their choices if they appeared to need help. Talking during the test was generally not permitted. The total testing time varied but usually was about one hour.

Scoring

Scoring of the answer sheets was done by using the answer keys provided in the test materials. These keys yield total "weighted item" raw scores on each of the 11 musical factors which together comprise the music preference test. Once these weighted raw scores were obtained from every S's answer sheet, the actual scoring was completed. These score values are the data used in the statistical treatments.

Although a total of approximately 300 patients were present on the wards, only 230 actually participated and submitted complete answer sheets. Of these 230 complete records, 190 were usable. Forty records had to be discarded for one reason or another, e.g., filling all the blanks out before the test actually began, copying the responses of one's neighbor, attending to the music intermittently, and so forth. Since the diagnostic labels were obtained from the hospital only for those patients who met the testing criterion and had usable answer sheets, the diagnoses of these 40 patients are not known.

TABLE 1
DESCRIPTIVE IDENTIFYING DATA ON SAMPLES

Sample Type	Sample Size	Age in Years		No. of Months in Hospital		Educational Level in Grades		No. of Re-admissions		% in Some Type Treatment	% of Voluntary Commitment	% of Non-white	No. Ss of Different Religions ^a			No. Ss of Given Marital Status ^b			
		Mean	Range	Mean	Range	Mean	Range	Mean	Range				P	C	O	M	S	D	O
<i>Males</i>																			
Organics	30	46.4	24-67	20.7	0-108	8.6	0-16	1.8	0-6	57	37	17	14	15	1	8	9	7	6
Schizophrenics	19	34.4	21-53	47.6	1-180	8.8	0-16	.8	0-5	74	0	42	10	77	2	6	10	2	1
Sociopathic Personalities	25	47.3	22-74	11.1	0-110	9.5	0-16	2.1	0-12	60	76	8	11	14	0	3	3	11	8
Paranoids	11	51.6	42-63	75.3	1-242	9.1	8-13	.7	0-4	73	0	18	3	7	1	7	1	2	1
Affectives	10	48.9	18-58	32.5	5-146	6.8	0-12	1.3	0-4	30	30	0	2	6	2	6	2	2	0
Psychoneurotics	9	31.8	15-63	92.3	2-255	8.3	6-11	.1	0-1	33	11	0	3	2	4	0	8	0	1
Total Abnormals	104	44.0	15-74	46.6	0-255	8.5	0-16	1.1	0-12	58	33	16	43	51	10	30	33	24	17
Normals	26	57.5	49-66									0	26	0	0				
<i>Females</i>																			
Organics	13	41.5	25-62	9.7	1-111	8.4	0-12	2.2	0-9	77	15	8	8	5	0	4	3	2	4
Schizophrenics	32	39.6	21-63	53.7	3-139	10.1	0-14	.8	0-6	62	3	31	14	14	4	10	9	4	9
Sociopathic Personalities	8	45.2	26-64	17.2	1-75	8.1	0-13	1.1	0-5	50	25	25	4	4	0	2	1	1	4
Paranoids	10	55.7	35-74	71.1	5-252			.4	0-2	40	10	20	4	5	1	4	3	2	1
Affectives	15	42.0	26-58	41.7	4-202	8.1	0-12	2.3	0-6	60	0	7	4	11	0	7	4	0	4
Psychoneurotics	8	45.5	16-62	21.9	2-75	8.8	7-16	4.1	0-9	75	50	0	5	3	0	3	2	2	1
Total Abnormals	86	43.3	16-74	35.9	1-252	8.7	0-16	1.8	0-9	62	12	19	39	42	5	30	22	11	23
Normals	34	55.2	45-66									0	34	0	0				

^a P = Protestant; C = Catholic; O = Other (Jewish, none, or not known).

^b M = Married; S = Single; D = Divorced; O = Other (separated, widowed, or not known).

Statistical Treatment of Results

The statistical treatment of the results involved three steps, each step designed to determine whether or not a certain variable or category had an influence on scores of music preference. First, an analysis of covariance was carried out on each of the 11 music factors involved in the test. This analysis was employed in order to demonstrate that a classification based on diagnosis and sex would show significant effects when effects attributable to age were removed. Second, on those factors for which the classification showed significant effects, a two-way analysis of variance (sex by categories) was carried out to ascertain whether significant score differences should be attributed to sex, diagnosis, and/or interaction. Finally, contingent upon the results of the analyses of variance, *t* tests were carried out to determine the location of statistically significant differences among the diagnostic groups. It was also planned that the analyses of variance and *t* tests should be carried out on scores adjusted for regression on age.

RESULTS

An analysis of covariance was run on each of the 11 music factors. Consequently, any effect of age on the music preference scores was statistically controlled for. Out of the 11 analyses of covariance, 7 were significant at the 10% level of significance agreed to justify further analysis.¹ Although this was the value of significance agreed to warrant further analysis, the probabilities associated with all but one of the *F* ratios were at the 5% level or less. For these 7 factors on which such a result was obtained it may be said that the variance among the means of the 14 groups, i.e., 7 categories by two sexes, was large enough to warrant further analysis. On the remaining 4 factors for which the null hypothesis was not rejected, no further analysis was undertaken.

An analysis of variance was run on each of the seven music factors which showed a significant analysis of covariance result. For pur-

poses of exploratory analysis, if the *F* yielded by the analysis of variance had a chance probability of .10 or less, it was to be considered a statistically significant value. Though this was the level of statistical significance adopted by the authors, the actual *F* ratios for the main effect of diagnosis had associated probabilities of .05 or less.

The seven analyses of variance had in common the result of a significant main effect due to diagnostic category. The findings showed that for a given factor this effect might be the only statistically significant one or that it occurred in addition to a main sex effect or an interaction effect. A main sex effect was found on Factors 1, 8, and 11, indicating that a consistently higher or lower score is obtained by a given sex regardless of diagnostic group membership. Using observed score values, an interaction effect was found on Factors 2 and 6, indicating a variable higher or lower score by a given sex depending on group type. (These interactions drop out, however, if adjusted scores are used.) Because sex effects were present on these factors, only same-sex comparisons were permitted in *t* tests of mean comparisons, as stated earlier. Although sex effects did not appear in the analysis of variance of Factors 5 and 10, the next part of the analysis proceeds, for the sake of clarity and consistency, by cautiously assuming that they might.

The *t* tests of mean comparisons were then run in order to determine just where the statistically significant differences due to sample type occur. Except for Factors 1 and 11, the means used in the comparisons were the observed means of the samples, not means after adjustment for age, since inspection of the regression coefficients indicated that adjusting for age was not necessary, save on Factors 1 and 11. In addition, no multiple comparison technique seemed justified since it typically involves a within-groups or pooled-variance expression, and this was not compatible with the heterogeneity of variance which exists in some of the comparisons. Therefore, independent estimates of the population variance were used in the calculation of every *t* ratio and the degrees of freedom associated with each directly calculated (see Walker & Lev, 1953, pp. 157-158). The main consequence

¹ Due to limitations of space, the summary tables of raw score means and variances (Tables A and B), the summary tables of the analyses of covariance and variance (Tables C and D, respectively), and the summary table of regression coefficients (Table E) are not presented in this report, but they have been deposited with the American Documentation Institute. Order Document No. 6331 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for 35 mm. microfilm or \$1.25 for 6 × 8 in. photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 2

THE *t* VALUES OF DIFFERENCES OF CLINICAL GROUPS FROM CONTROLS,
ONE ANOTHER, AND TOTAL ABNORMAL GROUP FOR MALES

Significance of Differences	Factors						
	1	2	5	6	8	10	11
Controls vs.							
Total Abnormals	1.1	4.2***	-1.1	-3.3***	5.1***	2.1**	-3.0***
Organics	0.7	4.8***	-1.8*	-1.8	3.6***	2.0**	-1.6
Sociopathic Personalities	0.3	2.7**	-1.6	-1.3	1.9	2.0	-1.0
Schizophrenics	1.4	2.4**	0.6	-2.8***	4.9***	0.5	-2.2**
Paranoids	-1.4	2.5**	-1.0	-4.3***	2.6**	2.4**	-0.3
Affectives	0.2	0.2	-0.8	-1.0	1.3	0.2	-1.0
Psychoneurotics	4.2	2.4**	0.8	-2.4**	5.9***	2.1**	-3.3***
Total Abnormals vs.							
Organics	-0.2	2.0**	-1.4	0.8	-0.4	0.1	0.8
Sociopathic Personalities	-0.7	-1.3	-1.1	2.0	-2.1**	0.6	1.2
Schizophrenics	0.9	0.5	1.9*	-0.7	1.7	-1.3	-0.7
Paranoids	-2.8	0.5	-0.6	3.0**	-0.3	1.3	0.9
Affectives	-0.5	-2.4**	-0.2	0.9	-0.7	-1.0	0.0
Psychoneurotics	4.3***	-0.4	1.5	-0.9	3.6***	1.0	2.5**
Organics vs.							
Sociopathic Personalities	-0.3	-2.3**	0.1	0.7	-1.2	0.4	0.4
Schizophrenics	0.8	-0.3	2.1**	-0.9	1.5	-1.0	-1.0
Paranoids	-2.1**	-0.3	0.1	-2.7**	0.0	1.1	0.5
Affectives	-0.3	-2.8**	0.4	0.3	-0.5	-1.0	-0.2
Psychoneurotics	3.7***	-1.4	1.8	-1.1	3.2**	0.9	-2.4**
Sociopathic Personalities vs.							
Schizophrenics	1.0	0.9	1.9*	-1.6	2.4**	-1.2	-1.2
Paranoids	-1.5	1.0	0.0	-3.4***	0.9	0.7	0.2
Affectives	0.0	-1.6	0.4	-0.1	0.2	-1.1	-0.5
Psychoneurotics	3.6***	0.4	1.7	-1.6	3.9***	2.3**	-2.6**
Schizophrenics vs.							
Paranoids	-2.6**	0.0	-1.4	-1.8	-1.2	1.7	1.1
Affectives	-0.8	-1.9	-1.2	1.0	-1.3	-0.2	0.4
Psychoneurotics	2.6**	-0.6	0.3	-0.4	2.0	1.5	-1.7
Paranoids vs.							
Affectives	1.2	-2.0	0.3	2.6**	-0.4	-1.5	-0.6
Psychoneurotics	5.1***	-0.6	1.4	1.1	2.8**	-0.1	-2.4**
Affectives vs.							
Psychoneurotics	3.0***	1.7	1.3	-1.2	2.6**	1.4	-1.7

Note.—A minus value indicates that the group preceding "vs." has the higher mean.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

of this procedure is in keeping with conservative presentation of results since the error involved is one of obtaining fewer significant *t* ratios than if a multiple comparison technique were employed.

In the following presentation of the *t* test

results it seemed advisable merely to set forward the findings at a descriptive level without any attempt at deeper psychological interpretation. This approach seems most appropriate considering the exploratory level and intent of this investigation. It also seems recommended

by the fact that, although the factors are well established empirically, their psychological interpretation is necessarily tentative.

Table 2 presents the significance of differences of the clinical groups from the controls, from one another, and from the total abnor-

mal group, for the males only. Table 3 presents the same results for the females. No *t* value having a probability greater than 5% was considered noteworthy except for such an occurrence on Music Factors 1 and 5. On these factors a probability of 10% was con-

TABLE 3
THE *t* VALUES OF DIFFERENCES OF CLINICAL GROUPS FROM CONTROLS,
ONE ANOTHER, AND TOTAL ABNORMAL GROUP FOR FEMALES

Significance of Differences	Factors						
	1	2	5	6	8	10	11
Controls vs.							
Total Abnormals	0.4	5.3***	-0.4	-5.6***	5.7***	3.0***	-1.8
Organics	0.2	1.6	-2.0*	-2.1	2.9***	2.5**	0.2
Sociopathic Personalities	-2.3**	2.6**	-1.0	-0.4	1.1	1.3	0.3
Schizophrenics	0.1	3.5***	-0.5	-6.1***	4.6***	3.0***	-3.0***
Paranoids	1.2	3.3***	2.1*	-1.7	1.8	2.8**	0.5
Affectives	0.1	3.8***	0.6	-4.2***	4.0***	0.3	-2.0
Psychoneurotics	0.3	1.5	-0.3	-2.4**	2.4**	1.1	0.8
Total Abnormals vs.							
Organics	-0.1	-1.2	-2.1**	0.4	0.0	0.7	1.4
Sociopathic Personalities	-3.5***	0.0	-0.9	3.5***	-1.5	-0.1	1.2
Schizophrenics	2.0**	-0.6	-0.2	-1.9	1.3	0.5	-2.3**
Paranoids	1.1	0.7	2.7**	1.6	-1.9	1.1	1.9
Affectives	-0.4	1.3	1.0	-1.3	0.6	-2.4**	-1.3
Psychoneurotics	0.0	0.0	-0.2	0.3	-0.1	0.2	1.7
Organics vs.							
Sociopathic Personalities	-1.8*	0.7	1.0	1.9	-1.1	-0.4	0.2
Schizophrenics	0.6	0.7	2.9***	-0.9	0.5	-0.3	-2.1**
Paranoids	0.8	0.2	4.7***	0.7	-1.2	0.4	0.2
Affectives	-0.1	1.6	2.6**	-1.0	0.3	-2.0	-1.7
Psychoneurotics	0.1	0.5	1.0	-0.1	-0.1	-0.1	0.6
Sociopathic Personalities vs.							
Schizophrenics	3.0**	-0.2	0.7	-3.9***	1.8	0.2	-1.8
Paranoids	2.9**	0.4	2.5**	-1.1	0.2	-0.7	0.0
Affectives	2.1*	0.9	0.3	-3.2	1.5	-1.4	-1.5
Psychoneurotics	2.1**	0.0	0.3	-1.8	1.0	0.2	0.4
Schizophrenics vs.							
Paranoids	0.4	0.8	2.3**	2.3**	-2.0	0.7	2.7**
Affectives	-0.9	1.3	0.8	-0.2	-0.2	-2.1**	0.0
Psychoneurotics	-0.5	0.2	0.0	0.9	-0.6	0.1	2.2**
Paranoids vs.							
Affectives	-1.0	0.4	-0.1	-1.9	1.7	-2.3**	-2.1**
Psychoneurotics	-0.8	-0.3	-1.6	-0.8	1.0	-0.3	0.4
Affectives vs.							
Psychoneurotics	0.2	-0.6	-0.6	0.9	-0.4	1.0	2.0

Note.—A minus value indicates that the group preceding "vs." has the higher mean.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

sidered statistically significant. This seemed in keeping with the 10% level of significance that was accepted in the analysis of variance on Factors 1 and 5. Three levels of significance are reported merely in order that a given reader may consider those values which suit his particular research requirements, not because the authors have three levels of confidence in the present results.

Tables 2 and 3 reveal that the control group differs significantly from each of the pathologic groups on at least one factor and often on five or six. Tables 2 and 3 also present the results which indicate the ability of the seven music factors to discriminate among the various syndrome groups themselves. Examination of Table 2 reveals that the male organics show a significant difference, when compared with the total abnormal group, on Factor 2. They are also distinguished from each of the various other diagnostic groups by a difference on some one factor. The female organics are more easily distinguished from the other groups since they differ most on only one factor. Their lower score on Factor 5 distinguishes them from every group except the neurotics and sociopathic personalities.

The male sociopathic personality group shows no difference from the affectives. However, they are distinguished from all the remaining groups by a significant difference(s). The female sociopathic personalities differ from the total abnormals, schizophrenics, and affectives by being higher on Factor 6, and from the paranoids by being lower on 5. In addition, they differ significantly from the total abnormal and all the other pathologic groups by scoring lower on Factor 1.

The male schizophrenic sample is distinguished from the organics by a higher score on 5, and also from the sociopathic personalities by a higher score on Factors 5 and 8. They are also characterized by scoring higher on Factor 1 than paranoids, and lower than psychoneurotics. The female schizophrenics show a greater number of distinguishing comparisons. Possibly Factor 11 will prove to be the most discriminatory of this reaction.

The lower score of the male paranoids on Factor 6 distinguishes them from all other groups except for the neurotics and schizophrenics. The psychoneurotics score higher,

while the schizophrenics score lower, on Factor 1 than the paranoids. They show no other difference from the schizophrenics, but have a lower score on Factor 8 than the neurotics. The female paranoid sample is principally characterized by being significantly higher on Factor 5 than most of the other groups.

A higher score on either or both Factors 1 and 8 distinguished the male psychoneurotics from the other samples. Moreover, a significantly lower score on Factor 11 is obtained by the neurotics except in comparison to the schizophrenics and affectives. The female psychoneurotics have no consistent significant differences from the other clinical groups on any given factor.

The male affective group differs significantly from organics, paranoids, and psychoneurotics on some factors, as have already been indicated. The female affective group differs from the organics by being higher on Factor 5, from the sociopathic personalities by being lower on Factor 6, higher on 1, and from both schizophrenics and paranoids by being lower on 10. They also differ from paranoids by being lower on 11.

To summarize these differences in terms of those which hold for both sexes—and these are of most value for the practicing clinician—we find:

1. Abnormals are higher on Factors 2, 8, and 10, and lower on 6 (and 11, if one goes to the 5% level). Among these, organics are higher only on Factors 2 and 8; neurotics on 8; and schizophrenics on 2, 6, 8, and 11.

2. Sociopaths differ from all other abnormals in being high on Factor 6.

3. Schizophrenics differ, notably from organics, in being high on Factor 5 and relatively low on 11.

4. Neurotics distinguish from most groups by being higher on Factor 1 and, less definitely, on Factor 8.

5. Paranoids are higher on Factor 11, and affectives are lower on 11, than most.

Since the component factor scores on these patterns are virtually independent, it will be seen that multiple correlations between the specific diagnostic category and the scores on the component factors in the different patterns could well be appreciable, but such finer cal-

ulation must be left until confirmation is available on other samples. Meanwhile it would seem that 7 of the 11 factors in this test are important for clinical diagnosis.

DISCUSSION

Although, as stated, the Music Preference Test factors are at a descriptive rather than an interpretive stage, some discussion of their associations with abnormality may be profitable.

The descriptions are at present based on 16 PF test correlations (Cattell & Anderson, 1953b) and for the seven factors involved here are as follows:

F1. Adjustment (vs. Frustrated Emotionality) which correlates negatively with ergic tension (Q₁).

F2. Parasympathetic Immunity vs. Sensitivity to Threat.

F3. Introspectiveness vs. Social Contact, having a substantial correlation with I factor (see Cattell, 1957).

F6. Anxiety vs. Paranoid Imperviousness, correlating with O factor (guilt proneness).

F8. Resilience vs. Withdrawn Schizothymia (H factor).

F10. Lack of Drive.

F11. Tenacity vs. Cyclothyme Relaxedness.

Some "face validity" is given to these descriptions from the *normal* range, when we move to the present *abnormal* sample, in several respects. First, paranoids are higher on "paranoid imperviousness," and schizophrenics differ from other abnormals in being high on Factor 5, introspectiveness. However, abnormals generally are marked by toughness, i.e., parasympathetic predominance and resilience, plus lack of drive and lack of guilt proneness. The neurotics score more "adjusted," Factor 1, than the psychotics, and more resilient. Affectives show more "cyclothyme relaxedness," Factor 11(-).

The last relation was also found in the Cattell & Anderson study (1953b), as well as the higher scores of schizophrenics on introspectiveness; but, except for schizophrenics and affectives the studies are not comparable in their groups. Although this much consistency exists, there are several other instances—such as the high score of sociopaths on anxiety, Factor 6, relative to other abnormal

groups—which are relatively inapt and suggest that we are only just beginning to grope our way to the true meaning of these music preference factors. Some similar oddities of label have been noted by Coroso (1958). This is not surprising considering how little has yet been done to relate the factors to other landmarks in personality measurement. Despite our inability to interpret, however, the fact remains that the test has on both occasions of experiment shown highly significant differences of abnormals from controls and among abnormal, psychiatric syndrome groups.

Age as a possible influencing variable in the present study with abnormals was shown to have no appreciable relation to music preference. This seems to be in agreement with studies which reported the same finding, although they used a different type of measurement and *normal* Ss (Keston & Pinto, 1955; Valentine, 1926). Theoretically, one might argue that such results are confirmatory evidence for music preferences being considered more a function of the temperament of the perceiver than a function of experience, transient needs, familiarity, and musical training, which vary with age. Perhaps temperament largely determines the music one likes and becomes familiar with.

Questions concerning the influence of such variables as race and marital status on the music preference factor scores are difficult to answer from the present samples. Insofar as an answer is possible concerning racial differences, analysis with nonwhites eliminated indicates that it makes no difference. However, these are variables requiring further investigation.

SUMMARY

An exploratory study in the general area of music preference and personality diagnosis was undertaken for the purpose of searching out music preference differences of various psychiatric samples.

The IPAT Music Preference Test of Personality was administered to a group of controls and to six "diagnostically" different groups of abnormal patients. The test was scored in standard fashion to yield scores on each of the 11 music factors comprising the

test. Previous researches in the area had shown age and/or sex might influence music preference. Consequently their possible influence was examined in the present study.

The results showed that age is not significantly related to music preference but that there are sex differences associated with music preference. Subsequently all music preference comparisons were made within the same sex. Seven out of the original 11 test factors yielded statistically significant *t* tests of comparisons. Several of the music factor scores obtained by various groups were considered in terms of their psychological meaning, as deduced from personality test correlations in normals, with the result that some designations seem appropriate and others in need of revision.

It was concluded that the Music Preference Test of IPAT holds promise as a diagnostic instrument, as indicated by the marked score differences between various groups. The conclusion was also made that the test is best employed by treating results for males and females separately. Because of limitations of the present investigation it is believed that further studies should be carried out in order to determine whether or not the relations identified in this research are replicable.

REFERENCES

- CATTELL, R. B. Principles of design in "projective" or misperception tests of personality. In H. H. Anderson & G. H. Anderson (Eds.), *Projective techniques*. New York: Prentice-Hall, 1951.
- CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book, 1957.
- CATTELL, R. B., & ANDERSON, J. C. *The I.P.A.T. Music Preference Test of Personality*. Champaign, Ill.: IPAT, 1953. (a)
- CATTELL, R. B., & ANDERSON, J. C. The measurement of personality and behavior disorders by the IPAT Music Preference Test. *J. appl. Psychol.*, 1953, 37, 446-454. (b)
- CATTELL, R. B., & SAUNDERS, D. R. Musical preferences and personality diagnosis. I. A factorization of 120 themes. *J. soc. Psychol.*, 1954, 39, 3-24.
- CATTELL, R. B., & SCHEIER, I. H. *The meaning and measurement of neuroticism and anxiety*. New York: Ronald Press, 1960.
- COROSO, JOAN. Relationship between the IPAT Music Preference Test of Personality and E, S, and T scores on the Guilford-Zimmerman temperament survey. Unpublished master's thesis, Univer. Connecticut, 1958.
- FAY, P. J., & MIDDLETON, W. C. Relationship between music talent and preference for different types of music. *J. educ. Psychol.*, 1941, 32, 573-585.
- KERR, W. A. Three studies in plant music. *Industr. Music News*, 1943, 1, 4-10.
- KESTON, M. J., & PINTO, I. M. Possible factors influencing musical preference. *J. genet. Psychol.*, 1955, 86, 101-113.
- PODOLSKY, E. (Ed.) *Music therapy*. New York: Philosophical Library, 1954.
- SCHULTZ, F. J. Testing listening power in music. *Music Supervisors' Nat. Conf. Yearb.*, 1933, 26, 306-312.
- SOPCHAK, A. L. Individual differences in reactions to different types of music, in relation to sex, mood and other variables. *Psychol. Monogr.*, 1955, 69 (11, Whole No. 396).
- SOPCHAK, A. L. Relation between MMPI scores and musical projective test scores. *J. clin. Psychol.*, 1957, 13, 165-168.
- VALENTINE, C. W. *An introduction to the experimental psychology of beauty*. New York: Dodge, 1926.
- WALKER, H. M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

(Received August 24, 1959)

PERSONALITY CORRELATES OF MANIFEST ANXIETY IN CHILDREN¹

LUCIANO L'ABATE

Washington University School of Medicine, St. Louis

Wirt and Broen (1956) have questioned whether any relationship exists between anxiety as measured by the Children's Manifest Anxiety Scale (CMAS) (Castaneda, McCandless, & Palermo, 1956; McCandless & Castaneda, 1956) and the concept of anxiety as used in the clinic. Such a lack of relationship may be confounded by sex differences in defending against anxiety. Although boys and girls may obtain similar scores on tests like the CMAS, they may use different defenses in coping with anxiety. Such a viewpoint assumes that anxiety is a variable, with deviations in behavior viewed as defenses against it. If the assumption is correct, as reviewers of emotional sex differences in children lead one to believe (Kuhlen, 1957, p. 274; Terman & Tyler, 1954, p. 1075), one would predict that different relationships would hold between the CMAS and various measures of adjustment in boys and girls.

Psychoanalytic writers (Bonaparte, 1953; Freud, 1946; Greenacre, 1948; Silverberg, 1952) have broadly suggested that defenses against anxiety may be divided into two major categories that are pertinent to the assumption underlying the present study, namely: girls handle anxiety through autoplasic defenses, while boys handle it through alloplastic defenses. One would predict that although boys and girls may report the same

amount of anxiety on a paper and pencil test like the CMAS, the relationships between their scores on such a test and various measures of adjustment may vary considerably. Measures that could be divided rationally into autoplasic and alloplastic defenses should also differentiate between boys and girls. Although maladjustment as a total measure may represent the combination of autoplasic and alloplastic defenses, girls should show a greater degree of relationship between manifest anxiety scores and autoplasic defenses such as daydreaming and feelings of inferiority, while boys should show a greater degree of relationship between manifest anxiety scores and alloplastic defenses, such as rebelliousness and general "nuisance value." No more specific predictions than the above could be made because of difficulties inherent in classifying measures of adjustment as being clearly either autoplasic or alloplastic.

The aims of the present investigation, therefore, are twofold, namely: to study whether sex differences exist in the management of anxiety as well as to specify the nature of such differences by correlating the CMAS with Rogers' Test of Personality Adjustment (Rogers, 1931a) and Rogers' Rating Scale of Adjustment (Rogers, 1931b).

PROCEDURE

Sample

Forty-nine boys and 47 girls in the fourth to eighth grades (9 to 13 years old) were administered the CMAS and the Rogers' Test of Personality Adjustment (1931a) in groups of 20 subjects (Ss) or less. There were 10 boys and 10 girls in each grade except for the fourth grade, which contained 9 boys and 7 girls. The Ss were selected from a public school population on the basis of their average or above-average IQ. The mean Otis Self-Scoring IQ for boys was 106.38, $SD = 8.22$, while the mean Otis Self-

¹ An earlier version of this paper was read at APA Convention, Cincinnati, Sept., 1959. The author acknowledges the cooperation of Pitt County Health Dept., and of the Board of Education, Greenville, N. C. The data of this investigation were collected by Laura Ciporin through the cooperation of Edna Baker, who also collected the teachers' ratings. This report was written during the author's tenure as USPHS postdoctoral fellow at the Institute for Psychosomatic and Psychiatric Research and Training of Michael Reese Hospital, Chicago, Illinois.

Scoring IQ for girls was 112.21, $SD = 9.33$. The background of the sample was strictly rural, lower to middle class economically. Their school was located in a town of 3,000 in eastern North Carolina.

To determine whether our North Carolina sample was similar to the original Iowa City sample in terms of manifest anxiety level and the number of lies (the score for lies is derived from 11 questions that buffer the CMAS), the means of our fourth, fifth, and sixth grade boys and girls were compared to Levitt's (1957) data. He compared the original rural sample to an urban (Chicago) sample, finding a much lower mean level of manifest anxiety in urban than in rural children (Table 1). Except for a significantly greater Lie score in boys, our North Carolina sample did not differ significantly from the original Iowa City sample in terms of anxiety level.

TABLE 1

ECOLOGICAL DIFFERENCES ON THE CHILDREN'S MANIFEST ANXIETY SCALE BETWEEN NORTH CAROLINA, IOWA CITY, AND CHICAGO SAMPLES FOR FOURTH, FIFTH, AND SIXTH GRADE BOYS AND GIRLS

	Anxiety					
	Boys			Girls		
	Mean	SD	N	Mean	SD	N
North Carolina	17.57	6.13	29	18.09	6.25	27
Iowa City	15.95	7.57	202	18.42	8.36	184
Chicago	13.59	7.56	161	12.87	8.23	161
	Lies					
	Boys			Girls		
	Mean	SD	N	Mean	SD	N
North Carolina	3.23	2.08	29	3.29	2.41	27
Iowa City	2.02	1.85	187	2.52	2.11	174
Chicago	3.92	2.42	161	4.29	2.30	161

Significance of Differences between Means

	Anxiety			
	Boys		Girls	
	t	p	t	p
North Carolina vs. Iowa City	1.10	ns	.02	ns
North Carolina vs. Chicago	2.69	.01	3.14	.01
	Lies			
	Boys		Girls	
	t	p	t	p
North Carolina vs. Iowa City	3.54	.001	1.71	ns
North Carolina vs. Chicago	1.31	ns	2.08	.05

Furthermore, our sample differed from the Chicago sample to the same extent as the Iowa City sample did (Table 1).

In view of the above, conclusions and generalizations based on our sample should apply only to comparable samples of rural children. Generalizations about the significance of our findings to urban children should be made with caution, because of the significant differences in anxiety level existing between urban and rural children. Although the ecological comparison was made with children between the fourth and sixth grades, it should be noted that our sample also included children from seventh and eighth grades, grades not included in the original standardization of the CMAS. The two higher grades were included because of the applicability of Rogers' Test to a wider age range than the CMAS.

Description of Criterion Measures

The Rogers' Test of Personality Adjustment is a paper and pencil questionnaire filled out by the child himself. It yields separate scores for personal inferiority, social maladjustment, familial maladjustment, and daydreaming. The four scores and their sum are so directed that the higher the numerical score, the greater the degree of maladjustment. Besides Rogers' original validation (1931a, 1931b), a recent study by Smith (1957) showed that the test differentiated satisfactorily among three groups of well adjusted, adjusted, and maladjusted boys in sixth grade. The means of our 49 boys on the Rogers' Test compare well by inspection with Smith's sample (Smith, 1957, Table 22). The sole exception may be the daydreaming score, which appears to be slightly higher in Smith's all urban sample.

The scores achieved on the Rogers' Test by our public school sample were compared to those of 38 children referred to a mental health clinic for behavior problems. The clinic children were of the same age span, intelligence, and socioeconomic and geographic background as the normal school children. There were more clinic boys (34) than clinic girls (4), the latter being an admittedly small sample. The mean scores of clinic children on Rogers' Test were significantly greater on all four subscales than the public school children, except for family maladjustment scores for boys and social maladjustment scores for girls (Table 2). There were no statistically significant differences between the means of clinic boys and clinic girls. Either in differentiating a pathological sample of children from normal children, or in comparing our sample with a normal sample of sixth grade boys, Rogers' Test of Personality Adjustment appears to possess sufficient validity to be used as an internal criterion measure for the CMAS.

As an external criterion measure, each child was rated by two different teachers on Rogers' Rating Scale of Adjustment (1931b). All ratings are so directed that a high score indicates a greater degree of maladjustment. For instance, a score of 1 on the scale of independence means "very independent, anxious to grow up," while a score of 5, which is the higher limit, means "completely dependent, wishes to

TABLE 2

SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF CLINIC AND SCHOOL CHILDREN ON ROGERS' TEST OF PERSONALITY ADJUSTMENT

	Clinic Boys (<i>N</i> = 34)		School Boys (<i>N</i> = 49)		<i>t</i>	<i>p</i>	Clinic Girls (<i>N</i> = 4)		School Girls (<i>N</i> = 47)		<i>U</i>	<i>p</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>			Mean	<i>SD</i>	Mean	<i>SD</i>		
Personal Inferiority	14.35	4.46	9.79	3.54	5.19	.001	17.25	5.54	10.72	4.63	2.05	.02
Social Maladjustment	17.44	5.01	14.59	4.63	2.64	.05	17.25	4.15	15.81	4.15	.50	ns
Family Maladjustment	9.12	3.50	7.77	3.40	1.75	ns	11.75	6.42	6.13	3.66	1.56	.06
Daydreaming	2.94	2.22	1.73	1.93	2.62	.05	4.75	2.85	2.79	2.18	1.54	.06
Total Personality Adjustment	43.85	8.09	34.00	6.52	6.08	.001	51.00	8.51	35.94	7.35	2.68	.005

remain infantile." A score of 1 on daydreaming means no daydreaming at all, while a score of 4 means "extreme daydreamer, of fantastic types." Unfortunately the same two teachers did not evaluate all the children so that commonly accepted measures of reliability like Pearson's *r* could not be applied.

Originally the Rating Scale had been constructed to be used by child guidance clinic workers. A few changes were made to adapt it for school teachers so that the teacher could rate a child according to what the child did in school in relation to teachers and peers rather than what he did at home in relation to parents and siblings. The total adjustment rating comprises seven subscales: daydreaming, rebelliousness, inferiority, affection, independence, conflict over sex matters, and the child's general emotional adjustment. The last two subscales were not considered individually because many teachers, all women, failed to rate children on conflict over sex matters and because the individual subrating of emotional adjustment would have added little information by itself. Each subscale consisted of one question with a choice of four or five statements of increasing degree of maladjustment. The inferiority ratings, however, are a composite of two questions, hence the greater numerical value for this subscale (Table 4). All of the

mean ratings reported are the sum of the ratings of both teachers for each subscale.

With two different teachers rating each child, the reliability of their ratings could only be estimated (Ebel's method, 1951). The teachers' ratings were on the whole surprisingly reliable (Table 3). With exceptions of ratings of inferiority in boys and daydreaming in girls, the high level of statistical significance reached by the ratings insures their usefulness as an external criterion for the CMAS.

RESULTS AND DISCUSSION

Although the mean IQ difference between boys and girls was statistically significant ($t = 3.32$, $p < .01$), no significant correlation between manifest anxiety level and intelligence for either sex was found ($r = .00$ in both sexes). There were no differences between the means of boys and girls on anxiety level and on the lie score of the CMAS or on personal inferiority, social maladjustment, and total personality maladjustment as measured by Rogers' Test. The mean score for family maladjustment was slightly higher in boys than in girls, while girls were slightly higher than boys on daydreaming. According to their ratings, teachers viewed boys as being consistently more maladjusted than girls in all aspects of Rogers' Scale (Table 4).

Highly significant correlations between subtests and the total score of Rogers' Tests as well as between teachers' subratings and the total ratings of maladjustment in both sexes were found. The correlation between the total score for personality maladjustment on Rogers' Test and the total sum of ratings on Rogers' Scale was statistically significant for boys and nearly so for girls.

TABLE 3

RELIABILITY FOR TEACHERS' RATINGS OF SCHOOL CHILDREN ON ROGERS' ADJUSTMENT RATING SCALE (Means and *SD*s in Table 4)

	Boys (<i>N</i> = 49)		Girls (<i>N</i> = 47)	
	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Total Adjustment	2.88	.005	4.13	.001
Daydreaming	1.76	.05	.84	ns
Rebelliousness	3.38	.001	3.44	.001
Inferiority	.83	ns	2.64	.005
Affection	2.66	.005	1.23	ns
Independence	2.72	.005	2.79	.005

TABLE 4

SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS FOR BOYS AND GIRLS ON THE CHILDREN'S MANIFEST ANXIETY SCALE, ROGERS' TEST, AND ROGERS' SCALE

	Boys (<i>N</i> = 49)		Girls (<i>N</i> = 47)		<i>t</i>	<i>p</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>		
Anxiety	17.45	5.86	18.38	6.77	.72	ns
Lies	3.12	1.92	2.74	2.42	.85	ns
Personal Inferiority	9.79	3.54	10.72	4.63	1.10	ns
Social Maladjustment	14.59	4.63	15.81	4.15	1.36	ns
Family Maladjustment	7.77	3.40	6.13	3.66	2.27	.05
Daydreaming	1.73	1.93	2.79	2.18	2.52	.05
Total Personality Adjustment	34.00	6.52	35.94	7.35	1.37	ns
Total Adjustment Rating	30.92	6.42	25.42	6.09	4.29	.001
Daydreaming Rating	3.31	1.04	2.70	.83	3.18	.01
Rebelliousness Rating	4.43	.91	3.79	.58	4.13	.01
Inferiority Rating	7.73	2.82	6.25	2.12	2.88	.01
Affection Rating	3.47	1.19	2.83	.87	3.00	.01
Independence Rating	3.94	1.19	3.21	1.20	2.98	.01

Correlations between anxiety and all other scores and ratings for boys were low, statistically nonsignificant, and consistently negative. Boys who reported more feelings of inferiority on Rogers' Test were also rated as being more maladjusted, more dependent, and as possessing more feelings of inferiority than boys not reporting the same feelings. Those boys who reported greater family maladjustment were also rated as being more maladjusted, as daydreaming more frequently, as possessing more inferiority feelings, and as lacking in affection. Boys who reported more overall maladjustment were also rated as feeling inferior and as lacking in affection. The two latter ratings correlated also with ratings of rebelliousness (Table 5).

In girls two positive and significant correlations were found between anxiety and the total personality maladjustment score as measured by Rogers' Test as well as between anxiety and teachers' ratings of independence. The more anxious and maladjusted the girl, the greater was the probability of her being rated as being more dependent. Girls who reported greater family maladjustment were also rated as being more rebellious. Girls who reported greater personality maladjustments were rated as being more dependent. Self-reports of daydreaming correlated with ratings of greater rebelliousness, more intense feelings of inferiority, and more dependence.

The latter ratings correlated with ratings of lack of affection and of greater feelings of inferiority (Table 6).

Although not all of the sex differences found were in the predicted direction, especially for boys, three additional analyses of the results were made to specify further the nature of such differences. In the first place, if manifest anxiety is related more to autoplasic rather than to alloplastic behavior, then girls should show a greater number of significant correlations between the measures of adjustment used. Girls do differ from boys on several correlations. A highly significant difference was found between correlations of anxiety and the total score on Rogers' Test ($z = 2.94$, $p < .01$). Girls showed significantly higher correlations than boys between anxiety and daydreaming scores ($z = 2.16$, $p < .05$); between ratings of maladjustment and of independence ($z = 2.27$, $p < .05$); and between ratings of inferiority and of independence ($z = 2.05$, $p < .05$). The only significant difference between correlations in which boys exhibited a higher correlation than girls was between ratings of rebelliousness and of lack of affection ($z = 2.25$, $p < .05$). The differences between correlations are in the expected direction, but could also be attributed to the teachers rating boys as being more maladjusted than girls.

In the second place, if girls do tend to

TABLE 5

CORRELATIONS BETWEEN CHILDREN'S MANIFEST ANXIETY SCALE, ROGERS' TEST, AND ROGERS' SCALE FOR 49 BOYS

	2	3	4	5	6	7	8	9	10	11	12	13
1. Anxiety	.00	-.13	-.12	.04	-.20	-.19	-.07	.12	-.03	-.22	-.04	-.03
2. Lies		-.04	.09	-.17	.06	-.02	.16	-.04	.31*	.04	.15	.11
3. Personal Inferiority			-.23	.13	-.04	.42**	.42**	.24	.11	.42**	.19	.42**
4. Social Maladjustment				-.07	.09	.57**	-.06	-.04	-.06	.09	-.01	-.20
5. Family Maladjustment					-.01	.52**	.41	.31*	.04	.34*	.28*	.19
6. Daydreaming						.32*	.26	.00	.09	.18	.22	.24
7. Total Personality Adjustment							.46**	.26	.07	.51**	.30*	.25
8. Total Adjustment Rating								.46**	.52**	.82**	.66**	.59**
9. Daydreaming Rating									.27	.17	.10	.23
10. Rebelliousness Rating										.33*	.48**	.06
11. Inferiority Rating											.58**	.46**
12. Affection Rating												.06
13. Independence Rating												

* $p < .05$.** $p < .01$.

TABLE 6

CORRELATIONS BETWEEN CHILDREN'S MANIFEST ANXIETY SCALE, ROGERS' TEST, AND ROGERS' SCALE FOR 47 GIRLS

	2	3	4	5	6	7	8	9	10	11	12	13
1. Anxiety	-.17	.11	.21	.24	.21	.40**	.26	.06	-.08	.24	.09	.30*
2. Lies		.08	-.13	-.02	.08	-.06	.05	.01	.16	-.03	.14	.03
3. Personal Inferiority			-.20	.30*	.04	.64**	.16	.11	.12	.10	.10	.23
4. Social Maladjustment				-.04	.13	.37**	.18	.11	-.22	.23	-.04	.24
5. Family Maladjustment					.07	.60**	.17	.28	.40**	.07	-.02	.17
6. Daydreaming						.41**	.02	-.20	-.03	-.06	.31*	.01
7. Total Personality Adjustment							.27	.16	.16	.20	.14	.34*
8. Total Adjustment Rating								.63**	.32*	.82**	.48**	.82**
9. Daydreaming Rating									.36*	.34*	.17	.46**
10. Rebelliousness Rating										.04	.05	.22
11. Inferiority Rating											.38**	.73**
12. Affection Rating												.24
13. Independence Rating												

* $p < .05$.** $p < .01$.

TABLE 7
ITEMS OF THE CHILDREN'S MANIFEST ANXIETY SCALE CHECKED POSITIVELY
MORE OFTEN BY EITHER SEX

Item Number	Description of Item	χ^2	p
Girls More than Boys			
11.	I am secretly afraid of a lot of things.	4.13	.05
28.	My feelings get hurt easily.	9.60	.001
32.	It is hard for me to go to sleep at night.	4.38	.05
39.	I am afraid of the dark.	11.98	.001
Boys More than Girls			
8.	I wish I could be very far away from here.	5.01	.025
10.	(Lie Scale) I would rather win than lose in a game.	10.09	.001
41.	(Lie Scale) I never get angry.	4.00	.05

autoplastic behavior more than boys, they should show it in the way they report about themselves. Indeed, an item analysis of the CMAS (Table 7) indicates that girls checked more often those items that show a greater tendency to admit feelings that boys would be leary of reporting, as Sarason and others have recently suggested (Sarnoff, Lighthall, Waite, Davidson, & Sarason, 1958).

In the third place, additional developmental trends may account for some of the sex differences. Boys showed variations in personal inferiority scores ($F = 2.59$, $p < .05$), with an increase up to sixth grade and a subsequent decline. Also among boys, daydreaming scores decreased significantly with age ($F = 3.16$, $p < .025$), reaching a mean of 1.00 in eighth grade. The decreases with age in daydreaming scores among girls was less pronounced ($F = 2.44$, $p < .10$). Variations in rebelliousness and daydreaming ratings with no clear declines or increases with age were also present among boys ($F = 3.04$, $p < .05$ and $F = 4.25$, $p < .01$, respectively). Girls showed variations in total adjustment ratings ($F = 2.64$, $p < .05$), daydreaming ratings ($F = 7.21$, $p < .001$), and ratings of affection ($F = 3.32$, $p < .01$), with no clear declines or increases with age. No other developmental trends were statistically significant in either sex.

In conclusion, although mean anxiety scores may be the same for boys and girls, yet each sex seems to cope with anxiety with differ-

ent defenses. Perhaps the construction of separate anxiety scales for boys and girls should be considered.

SUMMARY

The present study evaluates some aspects of sex differences between boys and girls in their handling of anxiety. Although children may obtain equal scores on a paper and pencil measure of anxiety like the Children's Manifest Anxiety Scale, the relationships between manifest anxiety scores and various measures of adjustment may be different in boys and girls. It was expected that girls would show positive correlations between manifest anxiety level and measures of autoplastic adjustment, like daydreaming and feelings of inferiority. Boys were expected to show positive correlations between manifest anxiety level and alloplastic defenses, like rebelliousness and greater "nuisance value" than girls. To investigate whether any sex differences existed according to these expectations, the Children's Manifest Anxiety Scale was correlated with Rogers' Test of Personality Adjustment and Rogers' Rating Scale of Adjustment.

Boys were consistently rated by their teachers as being more maladjusted than girls, but no positive correlation between manifest anxiety scores and various measures of adjustment used was found for boys. Positive and statistically significant correlations between manifest anxiety level and Rogers' total score

of personality adjustment and ratings of independence were found for girls. Additionally, sex differences between correlations, in an item analysis of the Children's Manifest Anxiety Scale, and in developmental trends within each measure of adjustment were presented and discussed briefly.

REFERENCES

BONAPARTE, MARIE. *Female sexuality*. New York: Int. Univer. Press, 1953.

CASTANEDA, A., McCANDLESS, B. R., & PALERMO, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, 16, 317-326.

EBEL, H. J. Estimation of the reliability of ratings. *Psychometrika*, 1951, 16, 407-424.

FREUD, ANNA. *The ego and the mechanisms of defense*. New York: Int. Univer. Press, 1946.

GREENACRE, PHILLIS. Anatomical structure and super-ego development. *Amer. J. Orthopsychiat.*, 1948, 18, 636-648.

KUHLEN, R. G. *The psychology of adolescent development*. New York: Harper, 1957.

LEVITT, E. E. Ecological differences in performance on the Children's Manifest Anxiety Scale. *Psychol. Rep.*, 1957, 3, 281-286.

McCANDLESS, B. R., & CASTANEDA, A. Anxiety in children, school achievement, and intelligence. *Child Develpm.*, 1956, 27, 379-382.

ROGERS, C. R. *Manual of directions: A test of personality adjustment*. New York: Assn. Press, 1931. (a)

ROGERS, C. R. Measuring personality adjustment in children nine to thirteen years of age. *Teach. Coll. Contr. Educ.*, 1931, No. 458. (b)

SARNOFF, I., LIGHTHALL, F., WAITE, R., DAVIDSON, K., & SARASON, S. A cross-cultural study of anxiety among American and English school children. *J. educ. Psychol.*, 1958, 49, 129-136.

SILVERBERG, W. V. *Childhood experience and personal destiny: A psychoanalytic theory of neurosis*. New York: Springer, 1952.

SMITH, L. M. The concurrent validity of six personality and adjustment tests for children. *Psychol. Monogr.*, 1957, 72(4, Whole No. 457).

TERMAN, L. M., & TYLER, LEONA E. Psychological sex differences. In L. Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1954, Pp. 1064-1114.

WIRT, R. D., & BROEN, W. E., JR. The relation of the Children's Manifest Anxiety Scale to the concept of anxiety as used in the clinic. *J. consult. Psychol.*, 1956, 20, 482.

(Received August 28, 1959)

A NEW SCALE OF SOCIAL DESIRABILITY INDEPENDENT OF PSYCHOPATHOLOGY

DOUGLAS P. CROWNE

Ohio State University

AND DAVID MARLOWE

College of Medicine, University of Kentucky

It has long been recognized that personality test scores are influenced by non-test-relevant response determinants. Wiggins and Rumrill (1959) distinguish three approaches to this problem. Briefly, interest in the problem of response distortion has been concerned with attempts at statistical correction for "faking good" or "faking bad" (Meehl & Hathaway, 1946), the analysis of response sets (Cronbach, 1946, 1950), and ratings of the social desirability of personality test items (Edwards, 1957). A further distinction can be made, however, which results in a somewhat different division of approaches to the question of response distortion. Common to both the Meehl and Hathaway corrections for faking good and faking bad and Cronbach's notion of response sets is an interest in the *test behavior* of the subject (*S*). By social desirability, on the other hand, Edwards primarily means the "scale value for any personality statement such that the scale value indicates the position of the statement on the social desirability continuum . . ." (1957, p. 3). Social desirability, thus, has been used to refer to a characteristic of *test items*, i.e., their scale position on a social desirability scale.

Whether the test behavior of *Ss* or the social desirability properties of items are the focus of interest, however, it now seems clear that underlying both these approaches is the concept of statistical deviance. In the construction of the MMPI *K* scale, for example, items were selected which differentiated between clinically normal persons producing abnormal test profiles and clinically abnormal individuals with abnormal test profiles, and between clinically abnormal persons with nor-

mal test profiles and abnormal *Ss* whose test records were abnormal. Keyed responses to the *K* scale items tend to be statistically deviant in the parent populations. Similarly, the development of the Edwards Social Desirability Scale (SDS) illustrates this procedure. Items were drawn from various MMPI scales (*F*, *L*, *K*, and the Manifest Anxiety Scale [Taylor, 1953]) and submitted to judges who categorized them as either socially desirable or socially undesirable. Only items on which there was unanimous agreement among the 10 judges were included in the SDS. It seems clear that the items in Edwards SDS would, of necessity, have extreme social desirability scale positions or, in other words, be statistically deviant.

Some unfortunate consequences follow from the strict use of the statistical deviance model in the development of social desirability scales. With items drawn from the MMPI, it is apparent that in addition to their scalability for social desirability the items may also be characterized by their content which, in a general sense, has pathological implications. When a social desirability scale constructed according to this procedure is then applied to a college student population, the meaning of high social desirability scores is not at all clear. When *Ss* given the Edwards SDS deny, for example, that their sleep is fitful and disturbed (Item 6) or that they worry quite a bit over possible misfortunes (Item 35), it cannot be determined whether these responses are attributable to social desirability or to a genuine absence of such symptoms. The probability of occurrence of the symptoms represented in MMPI items (and incorporated in the SDS)

in a college undergraduate population is undoubtedly low. Thus, the achievement of high SD scores may simply reflect the low frequency of pathological symptoms in this population and not the *needs* of Ss to present themselves in a favorable light. Of course, if one is only concerned with the properties of test items (their social desirability scalability), this is not a relevant issue. If, however, major importance is attached to the needs of Ss in psychometric situations and the influence of these needs on test responses, it is essential to be able to discriminate between the effects of item content and the needs of Ss to present themselves in a socially desirable (or undesirable) light.

In the present research, a social desirability scale was developed according to a different psychometric model, avoiding the ambiguities of the statistical deviance approach. Basic to this model is the sampling procedure employed in the selection of items from a defined universe. The population from which items were drawn is defined by behaviors which are culturally sanctioned and approved but which are improbable of occurrence. This will readily be recognized as the rationale underlying the Lie scale of the MMPI (Meehl & Hathaway, 1946). Items in the present scale, however, are probably less extreme than the Lie items.

METHOD

Development of Scale

A number of current personality inventories were consulted by the authors in order to devise a set of items for a new social desirability scale (M-C SDS). For inclusion in the scale, an item had to meet the criterion of cultural approval described above and was required to have minimal pathological or abnormal implications if responded to in either the socially desirable or undesirable directions. A set of 50 items meeting these criteria was submitted to 10 judges, both faculty members and graduate students in the Department of Psychology of Ohio State University, for social desirability ratings. The judges were instructed to score each item in the socially desirable direction from the point of view of college students, using true and false response categories. Unanimous agreement was obtained on 36 items and 90% agreement on 11 additional items. These 47 items constituted the preliminary form of the scale.

A major objective in the development of the M-C SDS was the elimination of pathology-relevant item content. To test this and for comparative purposes,

both the M-C SDS and the Edwards 39-item SDS (Edwards, 1957) were submitted to 10 additional judges, again including both faculty members and graduate students in the psychology department, for ratings of the degree of maladjustment implied by socially undesirable responses to the items. A 5-point scale, ranging from extremely well-adjusted (1) to extremely maladjusted (5), was employed for this purpose. The mean rating for all the items in the M-C SDS was 2.8, slightly below the midpoint of the scale (implies neither good nor poor adjustment). The mean rating for the Edwards SDS items was 3.9, indicating that the judges considered socially undesirable responses on this scale to be definitely indicative of maladjustment. The *t* test of the significance of the difference between these means is 15.27, which is significant well beyond the .0001 level.

The preliminary scale was then administered to 76 students in two introductory psychology courses, and an item analysis completed. There were 33 items that discriminated at the .05 level or better between high and low total scores. Of the 33 items, 18 are keyed true and 15 false, making a response set interpretation of scores highly improbable. These 33 items constitute the final form of the M-C SDS and are listed in Table 1 with the socially desirable response scoring indicated.

Reliability

The internal consistency coefficient for the final form of the scale, using Kuder-Richardson formula 20, is .88. This was computed on 39 Ss, 10 males and 29 females, who were enrolled in an undergraduate abnormal psychology class at Ohio State University. The mean age of this sample was 24.4 years, with a range of 19 to 46 years. Thirty-one of these Ss took the scale on two occasions separated by a month interval. A test-retest correlation of .89 was obtained.

Relationship to Edwards SD Scale

The correlation between the M-C SDS and the Edwards SDS is .35, which is significant at the .01 level. The sample on which this correlation is based included, in addition to the 39 abnormal psychology students, 81 students in a course on exceptional children. The correlation shows a general tendency for scores on the two tests to be associated.

In Table 2, the means and standard deviations of both SDSs are reported. The distribution of M-C SDS scores rather closely approximates a normal distribution, while negative skewness, consistent with previous findings (Edwards, 1957), is found for the Edwards SD distribution. It is interesting to compare the Edwards SD mean found in the present research with that originally reported. The means of 28.6 and 27.1 for males and females reported by Edwards are considerably lower than the value found in this study.

Correlations with Other Scales

A considerable portion of the research on social desirability has involved the correlation of SDSs with MMPI variables. To compare the present scale with

TABLE 1

THE MARLOWE-CROWNE SOCIAL DESIRABILITY SCALE

Personal Reaction Inventory

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is *true* or *false* as it pertains to you personally.

1. Before voting I thoroughly investigate the qualifications of all the candidates. (T)
2. I never hesitate to go out of my way to help someone in trouble. (T)
3. It is sometimes hard for me to go on with my work if I am not encouraged. (F)
4. I have never intensely disliked anyone. (T)
5. On occasion I have had doubts about my ability to succeed in life. (F)
6. I sometimes feel resentful when I don't get my way. (F)
7. I am always careful about my manner of dress. (T)
8. My table manners at home are as good as when I eat out in a restaurant. (T)
9. If I could get into a movie without paying and be sure I was not seen I would probably do it. (F)
10. On a few occasions, I have given up doing something because I thought too little of my ability. (F)
11. I like to gossip at times. (F)
12. There have been times when I felt like rebelling against people in authority even though I knew they were right. (F)
13. No matter who I'm talking to, I'm always a good listener. (T)
14. I can remember "playing sick" to get out of something. (F)
15. There have been occasions when I took advantage of someone. (F)
16. I'm always willing to admit it when I make a mistake. (T)
17. I always try to practice what I preach. (T)
18. I don't find it particularly difficult to get along with loud mouthed, obnoxious people. (T)
19. I sometimes try to get even rather than forgive and forget. (F)
20. When I don't know something I don't at all mind admitting it. (T)
21. I am always courteous, even to people who are disagreeable. (T)
22. At times I have really insisted on having things my own way. (F)
23. There have been occasions when I felt like smashing things. (F)
24. I would never think of letting someone else be punished for my wrongdoings. (T)
25. I never resent being asked to return a favor. (T)
26. I have never been irked when people expressed ideas very different from my own. (T)
27. I never make a long trip without checking the safety of my car. (T)
28. There have been times when I was quite jealous of the good fortune of others. (F)
29. I have almost never felt the urge to tell someone off. (T)
30. I am sometimes irritated by people who ask favors of me. (F)
31. I have never felt that I was punished without cause. (T)
32. I sometimes think when people have a misfortune they only got what they deserved. (F)
33. I have never deliberately said something that hurt someone's feelings. (T)

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE SOCIAL DESIRABILITY SCALES

Scale	N	Mean	SD
M-C SDS	120	13.72	5.78
Edwards SDS	120	31.83	5.06
(From Edwards, 1957)	84	28.6 Males	6.5
	108	27.1 Females	6.5

the Edwards SDS, Pearson product-moment correlations were computed between the two SDSs and the following MMPI and derived scales: *K*—Test-taking attitude; *L*—Lie; *F*—Validity and test-taking attitude; *Hs*—Hypochondriasis; *D*—Depression; *Hy*—Hysteria; *Pd*—Psychopathic Deviate; *Pa*—Paranoia; *Pt*—Psychasthenia; *Sc*—Schizophrenia; *Ma*—Manic; *Pr*—Prejudice (Gough, 1951); *St*—Status (Gough, 1948); *Es*—Ego Strength (Barron, 1953); *MAS*—Manifest Anxiety (Taylor, 1953); *A*—Anxiety (Welsh, 1956); *R*—Repression (Welsh, 1956).

The 39 Ss referred to above who served in the study were administered the M-C SDS, the 39-item Edwards SDS, and the MMPI in that order. The first two tests were given on the same day and the MMPI about a month later. Thirty-four Ss completed all of the tests and 37 of them completed all but the derived MMPI scales.

Table 3 presents the correlations between the M-C

TABLE 3

CORRELATIONS BETWEEN THE SOCIAL DESIRABILITY SCALES AND VARIOUS MMPI SCALES FOR 37 MALES AND FEMALES

MMPI Scales	M-C SDS	Edwards SDS
<i>K</i>	.40*	.65**
<i>L</i>	.54**	.22
<i>F</i>	-.36*	-.61**
<i>Hs</i>	-.30	-.62**
<i>D</i>	-.27	-.72**
<i>Hy</i>	.15	.09
<i>Pd</i>	-.41**	-.73**
<i>Pa</i>	.21	-.02
<i>Pt</i>	-.30	-.80**
<i>Sc</i>	-.40*	-.77**
<i>Ma</i>	-.24	-.42*
<i>Pr</i> ^a	-.27	-.58**
<i>St</i> ^a	.16	.14
<i>Es</i>	.17	.46**
<i>MAS</i> ^b	-.25	-.75**
<i>A</i> ^b	-.23	-.61**
<i>R</i> ^b	.28	.07

* Significant at the .05 level.

** Significant at the .01 level.

^a N = 36.

^b N = 34.

SDS and the Edwards SDS and the 17 MMPI validity, clinical, and derived scales. It is at once apparent that uniformly higher correlations obtain between the Edwards SDS and the various MMPI scales than between the M-C SDS and these MMPI variables. A general trend, which is consistent with previous research, is found in the positive correlations between the SDSs and the validity scales of the MMPI, and negative correlations with most of the clinical scales. Four clinical scales correlate highest with both SDSs, with the single exception of *D* which correlated $-.27$ with the M-C SDS: *Sc*, *Pd*, *Pt*, and *Hs*. Two of these four, *Sc* and *Pt*, are considered to be among the most "pathological" of the clinical scales.

DISCUSSION

The most important feature of the findings of this study is found in the marked differences overall in the magnitude of the correlations between the two SDSs and the MMPI. Consistently higher correlations were found between the Edwards SDS and the MMPI scales than were obtained between the M-C SDS and the MMPI scales. The high Edwards SDS-MMPI correlations, in general, confirm findings previously reported by Edwards (1957) and Fordyce (1956). Correlations between the Edwards SDS and the *Pt*, *Sc*, and *MAS* scales, in fact, approach the asymptotic value of the reliabilities of the separate tests. With correlations this high, it is necessary to raise the question of whether the Edwards SDS and these MMPI scales are not, in effect, functionally unitary. It would appear to be difficult to hold the view that SD scores and *MAS*, *Pt*, and *Sc* scores can be interpreted differently. More in accord with the evidence would be to attribute the covariance of the Edwards SDS and these MMPI scales to item similarity and to the "pathological" content of both sets of items. This would lead to an interpretation of the Edwards scale as a measure of the willingness to admit to certain symptoms of a "neurotic" nature or as a measure of general "neuroticism." But this does not enable one to discriminate between high SD scorers who genuinely do not have the symptoms represented in the SDS items from those Ss who conceal (consciously or unconsciously) their symptoms and whose responses are motivated by social desirability. To the extent, then, that the Edwards SDS measures social desirability, it does so in the very re-

stricted sense that high *SD* scores imply that it is bad or undesirable to have or admit to symptoms. Possibly, such attitudes have little generality and would not be related to other test behavior or social behavior. Sarason (1959) has also raised the question of the interpretation of Edwards SDS as an unconfounded measure of social desirability.

In the development of the M-C SDS, social desirability was defined more broadly to refer to the need of *Ss* to obtain approval by responding in a culturally appropriate and acceptable manner. This conception does not involve the acquiescence or denial by *S* of pathology. The significantly different maladjustment ratings obtained on the two SDSs support the hypothesis that the Edwards SDS involves the admission or denial of maladjustive symptoms and indicate that socially undesirable responses on the M-C SDS do not imply maladjustment.

The smaller correlations between the M-C SDS and the various MMPI scales would be predicted if one views social desirability as accounting for a fraction of the MMPI variance but not all or most of it. The problem of overlapping meanings is thereby avoided. Thus, it is submitted that the M-C SDS-MMPI correlations more accurately indicate the amount of MMPI scale variance which may be attributed to differences in the need to give socially desirable responses.

It may additionally be pointed out that the M-C SDS and the Edwards SDS differ considerably in the amount of content or item overlap with the various MMPI scales. The present scale contains one exact and four approximate replications of *L* items and one repetition of a *K* scale item. By contrast, the Edwards SDS, it will be recalled, was constructed from a heterogeneous pool of MMPI items and not inconsiderably overlaps with many MMPI validity and clinical scales. The two SDSs have no items in common.

Certain additional aspects of the present findings are worthy of note. Positive correlations are found for both SDSs with the *K* and *L* scales, on which high scores are generally interpreted to indicate "defensiveness" and the attempt by *S* to cast himself in a favorable light. The M-C SDS correlates much more highly with *L*, however, than does the Ed-

wards SDS. The negative correlations with the *F* scale are accounted for by the interpretation of a high *F* score as an indication of "plus getting." Regarding the clinical and derived scales, in general those MMPI scales on which a high score indicates maladjustment are negatively correlated with the SDSs. In part, the exceptions to this may be explained in terms of the distinction between subtle and obvious scoring on some of the clinical scales. Item subtlety, meaning the relative absence of social desirability implications, would account for the negligible correlations between the Edwards SDS and the *Hy* and *Pa* scales, for example. The fairly substantial correlations between the Edwards SDS and the *Pr* and *Es* scales may again be a function of similarity in general item content. In the judgment of the present authors, about half of the items in the *Es* scale would be classed as "pathological," while roughly a third of the *Pr* items would be so considered.

The positive correlation between the M-C SDS and the *Pa* scale, however, is an interesting possible exception. While this *r* falls short of significance, it might suggest that high SDS scores (implying in the present definition of the construct a high need for the approval of others) tend to be associated with concern or suspicion about the motives of others. Correlations between the Edwards SDS and Welsh's *A* and *R* scales have not, to the writers' knowledge, been previously reported. The Edwards SDS correlated highly, as one would expect, with the *A* scale but not at all with *R* which has a rather heterogeneous item content in terms of pathology. The M-C SDS does better in this case with an *r* of .28. This is in the predicted direction since all of the items on the *R* scale are keyed false. The M-C SDS correlation with the *A* scale is of the same magnitude as the correlation with the MAS and is consistent with expectation. It would appear from the correlations of the SDSs with *A* and MAS that the latter are approximately equivalent measures.

SUMMARY

In this research, an alternative model to Edwards' conception of social desirability was proposed. Basic to the present construct of

social desirability is the definition of a population of culturally acceptable and approved behaviors which are, at the same time, relatively unlikely to occur. Test items were drawn from this population in the development of a new social desirability scale, the Marlowe-Crowne Social Desirability Scale. This scale was correlated with 17 MMPI validity, clinical, and derived scales and the results compared with the correlations of the Edwards SDS with these MMPI variables. The very high correlations obtained with the Edwards scale cast doubt on the interpretation of this test as a measure of the influence of social desirability on test responses. The magnitude of the correlations of the new scale with the MMPI was considered to be more in accord with a definition of social desirability in terms of the need of subjects to respond in culturally sanctioned ways.

REFERENCES

- BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327-333.
- CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, 6, 475-494.
- CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, 10, 3-31.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, 20, 171-175.
- GOUGH, H. G. A new dimension of status: I. Development of a personality scale. *Amer. sociol. Rev.*, 1948, 13, 401-409.
- GOUGH, H. G. Studies of social intolerance: II. A personality scale for anti-semitism. *J. soc. Psychol.*, 1951, 33, 247-255.
- MEEHL, P. E., & HATHAWAY, S. R. The K factor as a suppressor variable in the MMPI. *J. appl. Psychol.*, 1946, 30, 525-564.
- SARASON, I. G. Relationships of measures of anxiety and experimental instructions to word association test performance. *J. abnorm. soc. Psychol.*, 1959, 59, 37-42.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- WELSH, G. S. Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.
- WIGGINS, J. S., & RUMRILL, C. Social desirability in the MMPI and Welsh's factor scales A and R. *J. consult. Psychol.*, 1959, 23, 100-106.

(Received August 29, 1959)

RELATIONSHIP OF FAMILY DISTURBANCE TO COGNITIVE DIFFICULTIES IN A LEARNING-PROBLEM CHILD¹

MICHAEL A. WALLACH,

Massachusetts Institute of Technology

DAVID N. ULRICH,

Darien Public Schools, Connecticut

AND MARGARET B. GRUNEBaum

Judge Baker Guidance Center, Boston, Massachusetts

Psychologists interested in the relationship of motivation to cognition have obtained some understanding of the ways in which cognitive impairment can result from emotional disturbance. Meanwhile, clinical inquiries have led to an expanding knowledge of the effects of parent-child relationships upon the emotional processes of the individual. Little has yet been done, however, to trace the sequence from family relationships, through the child's emotional dynamics, to his cognitive processes. The present paper is designed to help bridge these areas by tracing a causal linkage between disturbance in parent-child interactions and the child's cognitive impairment.

The boy in question, Robert Strong, is now 15. He has had weekly hours of therapy for four years and has seen a therapeutically-oriented tutor for one to three hours a week over two academic years. His mother has seen a social worker for weekly hours during this whole period, and his father has seen the same person for the past two years. Eleven when he first came to the clinic, Bobby had repeated the second grade and had had near-failing marks in each grade through the fifth. Yet at the time of referral the Wechsler-Bellevue Intelligence Scale for Children confirmed that

his intelligence was average. After three and one-half years of treatment he has made marked gains in areas other than school performance, but he has yet to obtain an average report card. Regularly administered school achievement tests have left no doubt of his trouble in acquiring the basic skills and absorbing knowledge. Although these tests indicate that he is now above grade level in reading skill, this gain has not been reflected so far in his school performance.

THE PARENTAL RELATIONSHIP

The immediate family consists of Mr. and Mrs. Strong, a couple in their early 40's, their daughter, Flora, who attends college and comes home only on weekends, and Robert. Mr. Strong works as a semiskilled laborer in a factory, and his wife teaches nursery school.

On the surface, Mrs. Strong would wish us to see her as calm. She seeks to control the expression of emotions by reason and by rigid rules for behavior taken mainly from her religion. Education and the development of reasoning powers have been of great importance to her, since they help protect her against emotions and also constitute a means of expressing her rivalry with men. It is not a coincidence that her husband, in turn, is a person who has tiptoed through life. At 43 he is boyish in appearance and mannerism. In the beginning of treatment he described himself as "meek and shy, a self-conscious man with rusty brains." He bows out completely from the intellectual realm, deferring all to his

¹ This research was supported in part by two grants from the United States Public Health Service. One of these is under the direction of Bessie M. Sperry, and the other, of Jerome S. Bruner, whom the authors wish to thank for encouragement and aid. We are indebted to Beatrice Reiner, whose two years of work with the mother contributed much to the case.

wife and stating that his only contribution is that of brawn. Like his wife, his major defense is that of denial. Were we to take his word, anger, resentment, and rivalry are non-existent in his makeup. Yet, in truth, they accumulate to such an extent that he has explosive outbursts which frighten him as well as the rest of the family. In contrast to his wife's position, Mr. Strong has looked upon education and the ability to reason as powers which he could not afford to possess lest they make him destructive.

In the parents' marriage, a balance was formed that met both their needs. By virtue of her husband's need for her, Mrs. Strong has been able to deny her rivalry with men, to maintain at the same time a superior position to men, and also to continue to see herself as a helping person. If he feels unable to carve the meat, answer the phone, do the family finances, then she must take over these duties to help him. For Mr. Strong, the choice of marital partner was part of a lifelong renunciation.

Consider now the children of this marriage. Flora, their first-born, has managed to grow without disturbing this balance. She achieved a good academic standing and thereby formed a team with her mother of the educated women in the family. The role that was assigned to Bobby, on the other hand, is best understood by considering what the birth of a male child meant to these parents. For quite different reasons, a son was a threat to both of them. To Mrs. Strong, a male child was a potential competitor, a potential danger if he grew to mature manhood. To Mr. Strong, Bobby was perceived as a sibling rival. It was as though the father feared that Mrs. Strong would not have enough to give to two. At the same time, he saw that Bobby could possibly excel over him in manliness. We can see that the role Bobby adopted—that of passivity, dependence, and underachievement—met the needs of both his parents in quieting their fears.

THE BOY'S EMOTIONAL DIFFICULTIES

The parents' everyday behavior continually suggested that people contain dangerous impulses and the world is a dangerous place in which to live. Bobby had listened to his

father's detailed accounts of crippling accidents at the factory. He had heard his mother's equally thorough descriptions of the motorcycle accidents that had killed his maternal grandfather and maimed his maternal grandmother long before Bobby was born. He had overheard, while lying frightened and alone in his crib, the sounds that accompanied his maternal grandmother's death from a heart attack and fall down the stairs. Events such as these could well reinforce fear and guilt about his own aggressive impulses, especially when his mother did not let him forget that the bigger he got, the more easily he could hurt someone. Bobby found his attempts at mastery were often received by his parents as a hostile effort to put someone else out of first place. School was actually the most important sphere of achievement in Bobby's life, and both his parents felt strong ambivalence about his success in this area. Accordingly, accomplishment in school became endowed for him with fear and guilt.

On one occasion when the therapist discussed Bobby's fears about achievement in school, Bobby made no verbal response but worked his hand across the desk in a walking movement. When asked about this, he said it reminded him of a man with a broken leg on crutches. The weekend after receiving his first really encouraging school report, Bobby broke his skis in a fall. When asked how his sister would like it if he got a good report card, he replied that she would have a fit. When asked how his tutor would like it, he replied that his tutor would hit the ceiling. Bobby did not realize until it was pointed out to him that these phrases implied any hostility. On another occasion, he brought in a test paper containing an A and absent-mindedly erased the A from the paper while talking about it.

Even when his daily work had improved, he continued getting poor report cards. The low grades could be attributed to various self-defeating maneuvers that will be discussed below. No matter how genuine his distress about failure, it was evident that Bobby possessed a clearly defined need to fail.

Bobby seemed to perceive his relationship to his mother and father as laden with conflict. It was as if Bobby and the rest of the family were continually engaged in playing

"king of the mountain," and one could get enough space only by shoving someone else out of it. When asked if his father went ice skating with him, he replied with a grin, "No, I was using his ice skates." When Bobby proudly appeared in a pair of size 9 shoes, the therapist asked him what his father had said about this new possession. Bobby promptly answered, "He said now he could wear my shoes." When his wishes and fears about displacing father from the house were interpreted to Bobby, he replied with a dreamy smile, "You take the living room and I'll take the kitchen . . . and you'll starve to death."

Denial seemed to be Bobby's main means of reacting to family conflicts. He could never discuss clearly any central aspect of his relationship to his parents. When asked what was going on at home, he would say only that father had been doing some carpentry, or they had been working in the garden. During a period when the parents were known to be going through an explosive upheaval in their relation to each other, even the most persistent prodding brought no comment from Bobby. After extensive work with him, it began to appear that a close parallel existed between the way he thought about his family and the way he thought about schoolwork. Rather than grasp the crucial variables of any problem in their important relationships to one another, for instance, Bobby would fasten upon a fragmentary assortment of often peripheral details.

THE BOY'S COGNITIVE PROCESSES

Schoolwork's Danger Potential

Bobby was highly sensitized to the danger potential of everything in his environment. No matter with what he was engaged, those attributes of the object or activity that might injure him were the ones of which he was most aware, and hence the ones to which he responded. During one session, for example, the tutor suggested a word game in which one person provides a key word, and then the other must write out, as quickly as possible, all the words that the key word suggests to him. When the tutor gave Bobby "danger" as the key word, he replied, "I can answer that one

in a single word," and wrote "everything." He then proceeded to describe a car as something that runs you over, a plane as something that crashes, machines in factories as things into which the unwary worker can fall and get chopped up, a door as something that can catch and squash your finger, and on without end. In the case of each such object, attributes which are not very important to most children were the critical ones for Bobby and served to control his thought about the object.

If Bobby viewed everyday objects like cars and doors as potential sources of harm, this was even more true of his feelings toward school and tutorial sessions. Once, for example, when tutor and tutee were alternately making up radio "commercials" and recording them on tape, Bobby gave a commercial for the "Boston Singing Club," telling how you can get singing lessons, and as a result you develop a pitch that will break a glass. Singing lessons, his metaphor here for the tutorial sessions, apparently were seen as making him capable of doing harm.

Consider also the time when he and his tutor were drawing a graph to represent the relation of temperature to pressure of a certain gas. After working on the graph for a while, Bobby announced that he was sure there was going to be an explosion of the gas due to the pressure, so the pressure would go down; and he promptly extrapolated a dotted line on the graph which veered sharply downward. The tutor assured him that the scientist had designed a very strong steel chamber for the gas so it could not possibly explode even if the pressure were increased; but Bobby claimed rather that the chamber was bound to explode, the scientist get hurt, and the gas pressure drop. He then sketched in further drastic rises and falls of the dotted graph line, put snowcaps on all the peaks of the graphs—saying they are snow-covered mountains—and drew a figure skiing down the mountain side. Immediately afterward he drew a tree squarely in front of the skier, so the latter had to crash into it. When the tutor noted that a crash was imminent, Bobby then drew in two ski tracks going around opposite sides of the tree so that the skier would have to split in two in order to follow them. Thus, if Bobby could not claim his schoolwork would lead to

death through exploding gases, he would make it lead to death by some other means. Bobby's mind seemed far indeed from the ostensible goal of working on the graph. He was totally occupied with assaying the injury potential of his paper-and-pencil activity.

Much of Bobby's cognitive energy was absorbed by the task of detecting and guarding himself against such imagined dangers as these. He showed an avoidance of academic material in general and also an avoidance of the concept of equivalence in particular. We shall consider each in turn.

Physical Avoidance of Schoolwork

There were two levels at which Bobby sought to avoid academic material in general: the level of physical activity and the level of thinking. First of all, Bobby had developed to a high degree strategies for physically getting himself out of, or not getting into, schoolwork situations. Several times, for instance, he managed to "forget" to bring his books to tutorial, too often for it to be mere chance and despite specific injunctions not to forget. If the books were brought, another kind of procrastination often ensued: hiding so the tutor could not find him. When finally engaged in some sort of schoolwork, he would be as inactive at it as possible—yawning and sprawling over the table as if he had not slept for days, acting as if the pencil were too heavy a weight to lift. And at every possible juncture, he would try to lure the tutor away from the topic of schoolwork by starting to talk about other things or looking around for any object that might provide a distraction. At school he sometimes would do the wrong homework or fail to hand it in.

All children, to be sure, would rather do other things than schoolwork. But most of them learn to engage in school activities nevertheless, and even to find pleasure in them. Bobby, on the other hand, showed a degree of insistence in avoiding academic activities which went far beyond playful truancy.

Avoidance of Abstraction and Anticipation in Thinking

When physical means failed, Bobby's avoidance of academic material went on at the level of thinking itself. We have already suggested

that Bobby strongly denied his knowledge of crucial family relationships. Confronted in school with the task of determining which were the important facts and how they were related, Bobby was often at a loss. It seemed as if he were trying to avoid the kind of thinking that might lead back to intolerable discoveries. One part of thinking has to do with *abstraction* activities: summarizing a potpourri of data in some terse and effective manner that is easier to retain and use than were the original facts. Such is the function of a formula or of an outline for a book. Another process in thinking concerns what we may call *anticipation* activities: predicting the consequences of present events by a kind of trial-and-error in imagination. The effect of avoiding such abstraction and anticipation activities will appear as a kind of "perceptual seduction" by the specific information made available to the senses.

Bobby, in his tutorial work, gave evidence of avoiding both aspects of thought. Consider first the avoidance of abstraction. With regard to verbal material, for instance, he often complained that he never knew what he should remember, in history or science, from the jumble of words before him. In his note-taking he would copy out nonessential details verbatim from the text, and often miss the whole point of a passage. His own writing, in turn, was filled with kinds of errors that indicated an inability to rise above specific words and take in the breadth of a whole sentence or thought. The particular words themselves, each separated from its neighbors, held his complete attention—as, for instance, in the following sentences he wrote on diabetes: "Diabetes is inability of the body to oxidize sugar is known as diabetes. diabetes is a disease in the body that insulin is scarce that causes diabetes." This statement was written at age 13.

Avoidance of reasonable anticipations in his cognitive work was just as clear as his avoidance of abstraction. When assembling a radio kit in tutorial one day, he came to a place where three items had to be secured with a single screw. Bobby screwed one part in, discovered a minute later that another piece had to go there too and redid the whole screw assembly, then discovered soon again that yet a third piece was required and had to undo

everything once more. Later he came to another identical place and went through this same procedure, and still later this happened again. Even though the three assemblies were exactly the same, he did not profit from his past experience and anticipate that several parts had to be held by the same screw when he came to each subsequent screw assembly. Knowledge was never transferred, as if the planning that would be involved was a burden he refused to assume.

Avoidance of Equivalence Relationships

Bobby's avoidance of the threatening topic of relationships among family members spread to the matter of relationships among numbers too. Bobby had much difficulty with the notion that various arithmetic expressions are equivalent to one another—something his arithmetic teachers had been demonstrating to him for some time now. Even after studying these equivalences for weeks, he still could not tell the tutor the correct decimal equivalent of the fraction $\frac{1}{2}$ or the percentage equivalent of that fraction. When forced to try an answer, he would guess at random or would respond in terms of the specifics of the $\frac{1}{2}$ by claiming, for example, that $\frac{1}{2}\%$ was the required percentage equivalent. In similar fashion, Bobby could not translate from one fraction to another. For example, he could not guess what simpler fraction would be the approximate equivalent of $\frac{250}{525}$. Again, after tutor and tutee had indicated on a line the locations of 0, $\frac{1}{4}$, $\frac{1}{2}$, and 1, Bobby could not tell where $\frac{2}{8}$ would be located on this line, even though $\frac{1}{4}$ appeared there already; and he could not find the position of $\frac{3}{4}$, even though $\frac{1}{2}$ already appeared. This boy was 13 years old at the time.

During one session, Bobby was marking off sixths at equal intervals by means of vertical lines. He finally got to $\frac{6}{6}$ and then went over one further equal interval and labeled it "1." He thus put down $\frac{6}{6}$ and 1 at different positions and was quite satisfied with this until the tutor questioned him about it. It was as if he feared having the $\frac{6}{6}$ and the 1 displace each other—and the only way he could avoid this was to imply that they were not equivalent but different. This latent fear could well derive from his conceiving of family relation-

ships as involving no more than one possible actor for each role. When he was forced to deal actively with number equivalences, in turn, we found that once again he did not believe that two terms could fulfill a common function. His modes of thinking about the family hence may have generalized to an abstract cognitive domain.

What happened when Bobby was forced explicitly to consider number relationships is illustrated by such episodes as the following. The tutor, for example, set up the equation, $\frac{3}{5} = .6$, and asked Bobby: "Does this mean that the $\frac{3}{5}$ gets lost somehow or destroyed when we put down the .6, or rather does it mean the .6 and $\frac{3}{5}$ are just different ways of talking about the same thing?" Bobby decidedly indicated that an arithmetic term got lost, destroyed, when you made something else equal to it, and held firmly to this view of the matter despite lengthy explanations by the tutor on many different occasions. The tutor, for instance, presented such examples as $\frac{3}{5} = \frac{6}{10}$, and had Bobby cancel through the $\frac{6}{10}$ and thereby reduce it back to $\frac{3}{5}$. Here too he thought the $\frac{3}{5}$ got destroyed when made equal to the $\frac{6}{10}$, and he did not feel his logic undermined when he found the $\frac{3}{5}$ recoverable again from the $\frac{6}{10}$. Even when he was finally able to grasp the idea in the case of one example, he was still quite incredulous when the tutor told him that this is also what went on in the other cases that were discussed.

Later the tutor demonstrated in yet another way that these two fractions really referred to the same thing, by illustrating each of them on a pie-cut circle graph of the same diameter: the segments representing $\frac{3}{5}$ and $\frac{6}{10}$ of each circle, hence looking exactly the same. Finding out in this way that the proportion was the same on both circle graphs appeared to cause Bobby the greatest surprise, and further demonstration of other arithmetic equivalences by the same method continually brought forth exclamations of amazement. All this at an age, 13, when work with arithmetic equivalences is second nature to most children.

When the tutor drove Bobby home after a session in which the meaning of arithmetic equivalence had been explored, Bobby was inordinately prone to see the environment as

injurious: expostulating at great length on how cars might smash up against the curb and hurt it, on how dangerous cars were in general, and then himself almost running into the path of an oncoming car while crossing the street. During another session on number equivalences, he squirmed, was uneasy, and eventually broke off from his work to describe newspaper accounts of a boy who got twisted up in a clothesline and hung himself by accident, and a girl who was climbing a tree when she fell into a Y formed by the branches and choked to death. During the work itself, Bobby had been punching holes viciously with a pencil into the side of a cardboard box.

With regard to numerical equivalence relationships, then, Bobby suffered an impairment in his capacity for what Piaget (1954) calls "conservation of identity." In these number realms, any different manifestation or appearance of something meant, for Bobby, a different thing, and could never mean the same thing. This trouble with the conservation of identity perhaps had its roots in his fear that equivalence in the family was impossible, that one of the aspirants had to suffer destruction.

This section has described the impact of Bobby's emotional difficulties on his cognitive activities. We have noted his sensitization to the possibility of danger in school activities, his physical and cognitive avoidance of school-related matters in general, and his avoidance of the concept of equivalence in particular. Our discussion has traced these cognitive impairments to two main kinds of causes: fear of school achievement and avoidance of knowledge of crucial family relationships.

CONCLUSION

We have sought to demonstrate how a long-standing disturbance in family relationships can result in cognitive difficulties for a child. The effort to avoid awareness of the family conflict has led to chronic interference with formal thought processes, while the meanings attached to school achievement have had a further inhibitory effect. Treatment over a long period has brought about a variety of marked changes in the parents and in the boy, and he seems to be meeting the other challenges of adolescence with a degree of success that could not have been predicted as recently as a year ago. In particular, he has overcome an initial strong shyness with respect to girls and now boasts a steady girl friend, he does very well in school sports and shop, and he has been making more friends among peers. However, his general thinking capacity still shows partial impairment, and his capacity for reproducing knowledge under the stress of a school setting is still markedly impaired. In retrospect, it appears that successful therapeutic intervention in a case of this kind requires a much closer collaboration of therapeutic and educational methods than has yet generally been achieved. Most instructive for psychology, perhaps, is the degree of both specific and general relationship found between family setting and the child's cognitive processes.

REFERENCE

- PIAGET, J. *The child's construction of reality*. New York: Basic Books, 1954.

(Received September 2, 1959)

CASE HISTORY CORRELATES OF THREE MMPI PROFILE TYPES¹

HAROLD GILBERSTADT AND JAN DUKER

University of Minnesota

In spite of the accumulation of research studies attesting to the superiority of the actuarial over the clinical approach in describing personality (Meehl, 1954), surprisingly little attention has been given to the task of compiling descriptions to be applied actuarially. As a result of his survey of this research, Meehl (1956) called for the development of "cookbook" descriptions. The purpose of the present paper is to report on a project designed to develop a source book of personality descriptions for narrowly specified MMPI profiles. It is hoped that the use of such descriptions will improve communication about patients, will make profile interpretation less dependent on the personal experience and skill of the individual clinician, and will help to standardize classification of personality for further research. The present paper reports data from a trait and symptom check list which was applied to the case records of three MMPI types.

METHOD

Population

Records were selected from the files for all patients admitted to the Psychiatric Service of the

¹The authors wish to express appreciation to Starke R. Hathaway and to Robert D. Wirt of the University of Minnesota for their helpful suggestions concerning the preparation of this paper.

Minneapolis Veterans Administration Hospital and examined by the psychology section during the years 1952-1957. Of approximately 3,150 such cases, about 2,300 were first admission. All patients were male veterans. First admission cases which met the following criteria were studied:

1. MMPI administered not more than 21 days before or 21 days after admission.
2. Age range 20 to 60 years.
3. Primary diagnosis not brain damage.
4. MMPI scale *T* scores: *L* = 60 or less, *F* = 85 or less, *K* = 70 or less.
5. Shipley Institute of Living Scale IQ estimate 105 or more.

Selection of Profile Types

Of several methods of profile analysis (Cronbach & Gleser, 1953; Lykken, 1956; Mosel & Roberts, 1954), it was decided that the contingency method employed by Meehl and Dahlstrom (Meehl, 1959) was best suited for the project. This method consisted of the empirical derivation of a set of rules which would specify the elevations and the interrelationships of the MMPI scales of a given type of profile and which would exclude profiles assumed to be not of the given type. The method permitted utilization of experience which the investigators had accumulated with respect to the MMPI profiles of the population being studied. On the basis of this experience, a set of specimen MMPI profiles was collected for study which was expected to be representative of frequently recurring personality types in the setting. Characteristics of three profile types will be reported in this paper; namely, MMPI codes 2-7-8 (*D-Pt-Sc*), 2-7 (*D-Pt*), and 2-7-4 (*D-Pt-Pd*). Starting from the specimen profiles, a set of con-

TABLE 1
MEAN MMPI *T* SCORES FOR THE THREE CODE TYPES

	Scale													
	<i>N</i>	<i>L</i>	<i>F</i>	<i>K</i>	<i>Hs</i>	<i>D</i>	<i>Hy</i>	<i>Pd</i>	<i>Mf</i>	<i>Pa</i>	<i>Pt</i>	<i>Sc</i>	<i>Ma</i>	<i>Si</i>
Code type														
2-7-8	22	48	72	49	73	99	74	77	70	70	98	99	53	77
2-7	13	50	55	55	68	94	74	62	58	60	84	62	47	58
2-7-4	27	48	62	50	66	90	71	83	58	64	83	68	60	60

TABLE 2
PERCENTAGE FREQUENCIES OF CHECKLIST ITEMS FOR GENERAL
ABNORMAL (GA) SAMPLE AND EACH PROFILE TYPE

Check List Item	GA	Greater than GA			Equal to or less than GA		
		2-7-8	2-7	2-7-4	2-7-8	2-7	2-7-4
Agitated	14		23		5		7
Anxiety	33	46	85**	56*			
Anxiety reaction							
diagnosis	25		76**	44	18		
Auditory hallucinations	4	14				0	4
Back pain	5		8	11	5		
Blindness, eye complaint	3	5	15				0
Blunted or in-							
appropriate affect	18	41*				0**	0**
Cardiac complaint	3	5	23*	4			
Chest pain	3		15	4	0		
Compulsive	6	14	23	7			
Conflict with parents	14				5	0*	0**
Conflict with siblings	11				0**	0*	0**
Conflict with wife	30			56*	14	15	
Confusion (Nonorganic)	11				9	0*	0**
Dependent	19		23	30	18		
Depression	43	87**	85**	52			
Depressive reaction							
diagnosis	8	18	23	30**			
Difficult concentration	2	28**	8				0
Divorced	12			41**	5	0*	
Dizziness	6		15		5		0*
Epigastric complaint	17				9	15	11
Evasive-defensive	22				0**	8	4**
Father alcoholic	9				5	0*	7
Father domineering	4	23*				0	0
Father strict	3	18*	8	4			
Fearful	10	32*					
Feelings of hostility	17				9	8	7
Feelings of sex						8	0**
inadequacy	8	14					
Financial difficulty	6		8	26**	5	0	4
Guilt	13	18	15	26			
Headache	12	18	30				
Heavy drinking	32			96**	5**		7
Hostile	10		23	15	9	0**	
Ideas of reference	11	32*					
Immature	17	28		30		0*	0**
Impulsive	6			11	0*	8	
Inadequacy feelings	8	37**	23	11		0	
Indecision	3	23**		7			
Inferiority feelings	5	37**	23			0	
Insomnia	11	14	46**	30*			4
Irritable	11		30	26			
Loss of appetite	7	18	38*	15	9		
Loss of interest	2	23**		4			
Married	58		100**			0	
Mother physically ill	4	14			14**		41
Neck pains	3	5	15	7		0	0
Nervousness	30	37	69*	48			
Numbness	6	14	8				
Obsessions	1	18**	30**				0*
Paranoid delusions	10						0
Paranoid trends	10	14			0**	0*	0**
						8	4

TABLE 2 (Continued)

Check List Item	GA	Greater than GA			Equal to or less than GA		
		2-7-8	2-7	2-7-4	2-7-8	2-7	2-7-4
Passive	8	18					
Poor work adjustment	43			45	23	8	4
Restless	13		23		5	23	
Rigid	3		15		0		11
Rumination	1	28**	15*	4			0
Schizoid	8	46**					
Schizophrenic						8	4
reaction diagnosis	16	50**					
Suicide attempt	6	9		15		0**	0**
Suicidal preoccupation	12	23				0	
Suspiciousness	9					0*	11
Tearfulness	7	14			9	0*	0**
Tension	21	41	8				
Tremor	14		69**	52**			4
Unmarried	30	82**		26	14	0*	
Weak, tired, fatigued	9	32*	38*			0**	7**
Weight loss	6	14					7
Wife pregnant or post	5		23	15	0	0	4
Withdrawn, introversive	16	51**					
Worrying	9		46**	19	9	15	4

* Difference significant at .05 level.

** Difference significant at .01 level.

tendency rules was formulated for each type which would narrowly specify the characteristics of the profile type.

The case records for each of the three types were studied in the following manner. First, all cases for the year 1956 whose MMPI profile fit the rules and the criteria outlined above were selected. These records were read. They consisted of psychiatric discharge summaries and social case history data. This first reading gave a subjective clinical impression of the patients in the group. These records were then re-read and placed on a five-point scale from poor to excellent for goodness of fit to the preliminary subjective clinical formulation. Rules were then revised to eliminate cases which did not seem to fit well. Next, all records from the additional years of 1952 and 1953 were read and rules were again refined to eliminate cases which seemed inappropriate. Finally, the records for the remaining three years (1954, 1955, and 1957) were read and cases which did not receive high ratings for the developing types were considered to be misses. The cases which were eliminated in the process of refining the rules are not reported here in detail, although some characteristics of them will be discussed below.

The checklist was based on one developed empirically by Cantor (1952) who tallied descriptive terms appearing in discharge summaries of psychiatric patients in the Minneapolis Veteran's Administration Hospital. For the present study, three judges independently checked the list for each case. When two of the three judges agreed, the item was included in the frequency counts.

RESULTS

Table 1 gives the mean *T* scores of the MMPI scales for each of the profile types.

Table 2 presents the incidence of the checklist items for the three code types and for patients in general (Rosen, 1952). Items which were of less than 10% frequency for at least one of the groups have been omitted from the table. Significances of the differences in item frequencies between the groups were computed from a nomograph for calculating the significance of the difference between two percentages (Lawshech and Baker, 1950).

2-7-8 Type (*N* = 22)

The following set of rules was developed to specify the 2-7-8 MMPI code type:

1. *D*, *Pt*, and *Sc* over *T* score 70.
2. *D* minus *Sc* less than 15 *T* scores.
3. *Pt* and *Sc* less than 20 *T* scores apart (unless *D* or *Pt* or *Sc* greater than 100 in which case discrepancy can be greater).
4. If *Pt* is peak, include only if all other scales are below 90 and *Pt* is not more than 5 *T* scores greater than *D*.
5. *Si* greater than *Ma*.
6. *Ma* below *T* score 70.
7. *D* minus *Hs* more than 10 *T* scores.

TABLE 3
COMPARISONS BETWEEN PROFILE TYPES ON CHECKLIST ITEMS

Check List Item	278 > 27	278 > 274	27 > 278	27 > 274	274 > 278	274 > 27
Anxiety			*			
Auditory hallucinations	*					
Blindness, eye complaint				*		
Blunted or inappropriate affect	**	**				
Chest Pain			*			
Conflict with wife					**	**
Depression		**		*		
Difficult concentration		**				
Dizziness				*		
Father domineering	**	**				
Fearful		*				
Feeling of sex inadequacy	*					
Financial difficulty					*	
Heavy drinking					**	**
Ideas of reference	**	**				
Impulsive					*	*
Inadequacy feelings		*				
Indecision	**					
Inferiority feelings		**				
Insomnia			*			
Loss of interest	**					
Mother physically ill	*	**				
Numbness		**				
Obsessions		**		**		
Poor Work Adjustment						
Quiet	*	**				*
Rigid			*	*		
Rumination		*				

* Difference significant at .05 level.
** Difference significant at .01 level.

TABLE 3 (Continued)

Check List Item	278 > 27	278 > 274	27 > 278	27 > 274	274 > 278	274 > 27
Schizoid	**	**				
Sexual difficulty	**					
Suicide attempt						*
Suicidal preoccupation	**					*
Suspiciousness		*				
Tremor	*					*
Weak-tired-fatigued		*		*		
Weight loss	*					
Wife pregnant or post			*		**	
Withdrawn-introversive	*	**				
Worrying			*			

8. *Pt* minus *Pd* more than 10 *T* scores.

9. *Pt* minus *Pa* more than 10 *T* scores.

Patients obtaining profiles of the 2-7-8 type were checked more frequently than patients-in-general on the following items which were significant at the .01 level: depression, difficult concentration, inadequacy feelings, indecision, inferiority feelings, loss of interest, obsessions, rumination, schizoid, schizophrenic reaction diagnosis, unmarried, weak-tired-fatigued, and withdrawn-introversive. Significantly different in frequency from patients-in-general at the .05 level were: blunted or inappropriate affect, father strict, fearful, and ideas of reference. Diagnostically, these patients appeared to fit the characterization of the group of pseudoneurotic schizophrenics described by Hoch and Polatin (1949).

2-7 Type (*N* = 13)

The following set of rules was developed to specify the 2-7 MMPI code type:

1. *D* and *Pt* over *T* score 70.
2. *D* greater than *Pt*.
3. *Pt* minus *Sc* at least 15 *T* scores.
4. *Si* below *T* score 70.
5. *Pd* and *Pa* below *T* score 80 unless *D* and/or *Pt* not over 85 in which instance *Pd* and *Pa* must be below 70.

Patients obtaining profiles of the 2-7 type were checked more frequently than patients-in-general on the following items which were significant at the .01 level: anxiety, anxiety reaction diagnosis, depression, insomnia, married, obsessions, tension, and worrying. Significantly different in frequency from patients-in-general at the .05 level were: cardiac complaint, loss of appetite, nervousness, rumination, and weak-tired-fatigued. Diagnostically these patients seemed to fit the description of the anxiety reactions or obsessive-compulsive personalities.

2-7-4 Type (*N* = 27)

The following rules were developed to specify the 2-7-4 type:

1. *D*, *Pd*, and *Pt* above *T* score 70.
2. *Pt* minus *Sc* at least 5 *T* scores.
3. *Pt* minus *D* 10 *T* scores or less.
4. *D* minus *Pd* 20 *T* scores or less.
5. Among scales *D*, *Pd*, and *Pt*, two scores below *T* score 100.
6. If *Pd* is peak, include only if *Pd* does not exceed *D* or *Pt* by more than 10 *T* scores.
7. *Si* below *T* score 70.
8. *Ma* less than *Sc*.
9. *Ma* above *T* score 40.
10. If *Ma* is below *T* score 50 include only if *Si* does not exceed *Ma* by more than 15 *T* scores.

Patients obtaining profiles of the 2-7-4 type were checked more frequently than patients-in-general on the following items which were significant at the .01 level: depressive reaction diagnosis, financial difficulty, heavy drinking, and tension. Significantly different from patients-in-general at the .05 level were: anxiety, conflict with wife, and insomnia. Diagnostically these patients appeared to fit the stereotype of the oral-dependent alcoholics.

Items which were less frequent for any profile type than for patients-in-general can also be found in Table 2. It will be noted that patients-in-general tend to be described more frequently as evasive and defensive than either the 2-7-8 or the 2-7-4 profile types.

Table 3 shows the check list items which differed significantly between the various combinations of profile types. These differences occur despite the fact that all three profiles are of the same two-point code type.

DISCUSSION

The rules for specifying profiles have been presented in detail. With experience, deviations from the specified configurations are interpretable. For example, 2-7-8s with higher 4 (*Pd*) than permitted by the rules tended to have the general characteristics of the type but also tended to have histories of acting-out which frequently included police contacts and penitentiary sentences. With 6 (*Pa*) higher than permitted by the rules, an increase in hostility, in suspiciousness and, particularly, evidence of overt psychosis was found. In short, inferences which could be made from single scale or pair-wise scale deviations from the basic configuration could be used to modify the interpretation of the profile. Similarly, although age and intelligence were restricted in deriving the descriptions, cases which deviated on these variables tended to fit less well but usually were not complete misses. The rules have been devised to select cases which are highly purified for research purposes but in many instances the cutting point necessarily was somewhat arbitrary and resulted in the exclusion of cases which would have fit fairly well. For example, for the 2-7-8 profile type, the rule of

D minus *Sc* of 15 *T* scores or less (Rule 2) was set to exclude depressive reactions even though some pseudoneurotic schizophrenics may be missed by rigid application of this rule.

It should be noted on the one hand that the three MMPI profile types reported would be expected to show many common traits relating to anxiety and depression since they were all of the 2-7 code type; on the other hand, there are many clear-cut differences despite the fact that they are all of the same two-point code type. For the sake of brevity more extensive clinical narrative descriptions which elaborate on the constellations of traits and contain inferences derived from them have been omitted. However, such descriptions have been developed for clinical use.

Previous research (Gilberstadt, 1952) utilizing the CC index of MMPI profile similarity devised by Hathaway and Meehl (Sutton, 1952; Williams, 1952) served as a pilot study for the present research. The study of the 2-7-8 profile type is essentially a cross-validation of earlier findings. In both studies the cases obtaining profiles of this type were highly similar to the syndrome of pseudoneurotic schizophrenia described by Hoch and Polatin (1949).

Further validation of the profile types by the Q sort technique is planned. This approach, when applied with descriptions derived from MMPI *Atlas* (Hathaway and Meehl, 1951) case histories, has already been shown by Duker (1958) to be fruitful.

Current research includes the most frequent MMPI code types and extends in scope to construct validation with other tests.

SUMMARY

Contingencies for specifying three frequently occurring types of MMPI profiles of the *D-Pt* variety have been developed from a large pool of Veteran's Administration general hospital psychiatric cases. These include a *D-Pt-Sc* pseudoneurotic schizophrenic type, a *D-Pt* obsessive-compulsive type, and a *D-Pt-Pd* alcoholic type. Check list data from social histories and psychiatric discharge summaries are presented for each of the types. These profile types can be used as a source

of reference for profile interpretation or as a means of classification for research purposes.

REFERENCES

- CANTOR, J. M. Syndromes found in a psychiatric population selected for certain MMPI code endings. Unpublished doctoral dissertation, Univer. Minnesota, 1952.
- CRONBACH, L. J., & GLESER, GOLDINE. Assessing similarity between profiles. *Psychol. Bull.*, 1953, 50, 456-473.
- DUKER, JAN. The utility of the MMPI *Atlas* in the derivation of personality descriptions. Unpublished doctoral dissertation, Univer. Minnesota, 1958.
- GILBERSTADT, H. An exploratory investigation of the Hathaway-Meehl method of MMPI profile analysis with psychiatric clinical data. Unpublished doctoral dissertation, Univer. Minnesota, 1952.
- HATHAWAY, S. R., & MEEHL, P. E. *An atlas for the clinical use of the MMPI*. Minneapolis: Univer. Minnesota Press, 1951.
- HOCH, P., & POLATIN, P. Pseudoneurotic forms of schizophrenia. *Psychiat. Quart.*, 1949, 23, 248-276.
- LAWSHECH, C. H., & BAKER, P. C. Three aides in the evaluation of the significance of the difference between two percentages. *Educ. psychol. Measmt.*, 1950, 10, 263-270.
- LYKKEN, D. T. A method of actuarial pattern analysis. *Psychol. Bull.*, 1956, 53, 102-107.
- MEEHL, P. E. *Clinical vs. statistical prediction*. Minneapolis: Univer. Minnesota Press, 1954.
- MEEHL, P. E. Wanted—a good cookbook. *Amer. Psychologist*, 1956, 11, 263-272.
- MEEHL, P. E. A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. *J. counsel. Psychol.*, 1959, 6, 102-109.
- MOSEL, J. N., & ROBERTS, JUNE. The comparability of measures of profile similarity: An empirical study. *J. counsel. Psychol.*, 1954, 18, 61-66.
- ROSEN, A. Development of some new MMPI scales for differentiation of psychiatric syndromes within an abnormal population. Unpublished doctoral dissertation, Univer. Minnesota, 1952.
- SUTTON, M. L. Profile patterning and descriptive correlates of patients having low scores on scale 9 of the MMPI. Unpublished doctoral dissertation, Univer. Minnesota, 1952.
- WILLIAMS, H. L. Differential effects of focal brain damage on the MMPI. Unpublished doctoral dissertation, Univer. Minnesota, 1952.

(Received September 14, 1959)

THE EFFECTIVENESS OF THE BENDER-GESTALT IN DIFFERENTIATING A FLIGHT GROUP FROM AN AGGRESSIVE GROUP OF ADOLESCENTS

LOREN V. COROTTO AND ROBERT H. CURNUTT

Napa State Hospital, Imola, California

There is some evidence to suggest that the Bender-Gestalt measures some aspect of ego strength. Pascal and Suttell (1951) developed an objective scoring system for the B-G that consists of tabulating certain deviations in the reproductions of the designs. These authors believe that "if . . . we limit our definition of ego functioning to the ability to reproduce faithfully the B-G drawings as presented, then we may say that ego strength lies on a continuum from very low to very high B-G scores" (p. 9). Therefore, the lower the score obtained on this quantitative continuum the greater is the individual's ego strength. One of the authors in another study was unable to relate $F+\%$ on the Rorschach with a Z score on the B-G (Curnutt & Lewis, 1954). In this study, however, an attempt will be made to follow up those results and, at least by inference, to see if it can be determined whether the B-G does measure some aspect of ego strength when used with behavioral criterion.

It is frequently noted in clinical practice that adolescents who tend to react aggressively to their environment often produce better B-G records than do adolescents who tend to utilize escape or flight behavior. It is generally assumed in the literature that people who utilize aggressive behavior tend to have more ego strength than individuals who tend to rely upon flight behavior. If the B-G does measure an aspect of ego strength, it should tend to differentiate between aggressive individuals and individuals who utilize flight. It is precisely this assumption which will be dealt with in this study.

The hypothesis to be tested may be stated as follows: The Pascal and Suttell scoring system of the B-G test will not significantly

differentiate adolescents who react primarily with flight behavior in contrast to those who react primarily with aggressive behavior.

PROCEDURE

The records of 92 adolescent patients who had been hospitalized in a children's psychiatric unit were withdrawn from the files. The patients in the aggressive group and in the flight group were matched individually on the basis of age, sex, and education. This was done since it is already known that both age and education are related to B-G scores (Bender, 1938; Pascal & Suttell, 1951). All individuals who had either a diagnosis or a known history of psychosis, mental deficiency, or organic brain damage were excluded.

The flight group was composed of the adolescents who had a record of two or more runaways from home since the age of 10 but with no resultant aggressive behavior. The aggressive group was defined as those subjects who had a record of two or more physically aggressive acts: e.g., destruction of property, fighting, or physical assault, but without any runaway behavior.

The selected sample was independently matched by one of the authors prior to obtaining the B-G protocols. The other author was subsequently presented with the randomized and codified protocols and asked to score them in the blind.

The sample utilized in the present study consisted of 27 matched pairs of males and 19 matched pairs of females. The ages of the sample ranged from 11 to 18. The Sign Test was used to analyze the obtained results.

RESULTS AND DISCUSSION

The analysis of the 46 pairs of adolescents failed to yield a significant difference, thus sustaining the null hypothesis. However, when the paired individuals were examined on a sex basis, significant differences were obtained: i.e., the male scores had tended to cancel out the female scores. Hence the results can be explained purely on the basis of sex differences. In view of this finding, an

analysis restricted to the 27 individually matched pairs of males in the original sample was undertaken with the result that the B-G did significantly differentiate between the aggressive group and the flight group. The difference was significant at the .001 level when applying a two-tailed test of significance. In other words, the aggressive males had lower B-G scores, i.e., more ego strength as operationally defined by Pascal and Suttell, while the flight males had the higher B-G scores.

When the 19 individually matched pairs of females in the original sample were examined, an inverse relationship to that of the males was obtained. The difference was significant at the .001 level. In contrast to the male sample, the female flight group was found to have the lower B-G scores (more ego strength) while the aggressive female group obtained the higher B-G scores. The marked reversal of the behavioral correlates associated with the lower B-G scores by the female sample in comparison to the male sample would seem to emphasize caution in attempts to relate behavior to B-G scores without thoroughly allowing for sex differences. Since Pascal and Suttell (1951) were unable to demonstrate a significant sex difference in the standardization of their scoring system, the possibility is raised that social and cultural influences may affect the behavioral expression of that aspect of ego strength tapped by the B-G.

These findings may have implications for the treatment of adolescents whose behavioral symptoms primarily revolve around aggression and flight. Since the results indicate that boys who utilize flight reactions have relatively less ego strength than do aggressive boys, then possibly the therapist might wish to direct his efforts more towards supporting

and developing the ego of the individual. In the aggressive group of boys, however, the aims of the therapist might be different. In this case, the therapist's efforts might be directed more towards channeling unacceptable behavior into more socially acceptable modes of response. Similar considerations would seem applicable to a female population.

SUMMARY

In order to explore the effectiveness of the B-G in differentiating a group of adolescents who react primarily in an aggressive manner from a group which reacts primarily in a flight manner, 46 pairs of adolescents were matched by age, sex, and education. When the groups were segregated on a sex basis, significant differences (.001) were obtained. Adolescent girls who utilize primarily flight behavior tend to have lower B-G scores (more ego strength as operationally defined by Pascal & Suttell, 1951) than do adolescent girls who react primarily with aggressive behavior. Adolescent boys who tend primarily to react with aggressive behavior, however, tend to have lower B-G scores (more ego strength) than do adolescent boys whose primary behavior reactions are characterized by running away. Some implications of these findings are discussed.

REFERENCES

- BENDER, L. A. A visual motor Gestalt test and its clinical use. *Amer. Orthopsychiat. Ass. res. Monogr.*, 1938. No. 3.
- CURNUTT, R. H., & LEWIS, W. B. The relationship between Z Scores on Bender-Gestalt and $F + \%$ on the Rorschach. *J. clin. Psychol.*, 1954, 10, 96-97.
- PASCAL, G. R., & TUTTELL, B. *The Bender-Gestalt test*. New York: Grune & Stratton, 1951.

(Received August 4, 1959)

BRIEF REPORTS

THE EPPS PROFILE STABILITY COEFFICIENT

DONALD W. FISKE, KENNETH HOWARD, AND WILLIAM RECHENBERG

University of Chicago

An unusual and almost uninvestigated feature of the Edwards Personal Preference Schedule (Edwards, 1957) is the Profile Stability Coefficient, the correlation (over needs) of the two sets of part scores for each S. This correlation was obtained for each of 135 college students. While Edwards uses r as the stability coefficient, we used ρ , since the hierarchy of the needs seems of primary concern. Edwards' distribution and ours were highly congruent, even to a second mode at .85-.89. Our mean, obtained via a rho-squared transformation, was .73, while his was .74.

Inspection indicates that the stability coefficient is in part a reflection of the relative scatter or differentiation of the S's profile of needs. It is also associated with separate scale scores. (In these analyses we used the square of ρ to reduce the skewness of the distribution of stability indices). For EPPS variables, the stability score correlated .37 with n Autonomy, .30 with n Exhibition, .23 with the Consistency score, and .17 with both n Achievement and n Dominance; it also correlated -.28 with n Abasement, -.27 with n Order, and -.21 with n Deference. (.22 is significant at the .01 level, .17 at the .05).

Among 19 other test scores, it correlated .22 with Dominant and .18 with Active on the Thurstone Temperament Schedule. It also correlated .19 with Expressed Control and -.22 with Wanted Control scores from the FIRO-B Scale (Schutz, 1958, pp. 57-65). (No association was found with two tests of intelligence). These correlations are consistent with those for the EPPS

scores. Consistency of a need profile on two groups of items appears to be related to active strivings for superior and controlling status.

For the same Ss, the relationship between positive self-ratings and intra-individual variability was negative for each of six traits, although no dependable relationships were found between such variability and the same set of test scores (van der Veen & Fiske, 1960). However, the temporal variability of mean peer ratings on assertiveness and number of useful ideas contributed to a group problem solving session was negatively related to self-descriptions of assertiveness and sociability (Fiske, 1960). These several findings suggest that individuals who describe themselves in positive and active terms may be less variable in both their test responses and their behavior.

REFERENCES

- EDWARDS, A. L. *Manual for The Edwards Personal Preference Schedule*. (Rev. ed.) New York: Psychological Corp., 1957.
- FISKE, D. W. Variability among peer ratings in different situations. *Educ. psychol. Measmt.*, 1960, 20, in press.
- SCHUTZ, W. C. *FIRO: A three-dimensional theory of interpersonal behavior*. New York: Rinehart, 1958.
- VAN DER VEEN, F., & FISKE, D. W. Variability among self-ratings in different situations. *Educ. psychol. Measmt.*, 1960, 20, 83-93.

(Received September 25, 1959)

MANIFEST NEEDS AND MANIFEST ANXIETY¹

CARL D. WILLIAMS, ROBERT B. TALLARICO, AND JAMES T. TEDESCHI

University of Miami

The present study was designed to contribute additional empirical properties of "anxious" and "nonanxious" groups on the Taylor Manifest Anxiety Scale (MAS) by comparing their manifest needs as measured by the Edwards Personal Preference Schedule (EPPS). The MAS was given to 543 college students, and extreme groups of 65 persons each were identified. The EPPS was given to the 38 MAS high scorers and the 39 MAS low scorers who were willing to participate upon invitation. There were 23 male and 15 female high scorers and 19 male and 20 female low scorers. The MAS scores of the anxious group ranged from 21 to 34. The means for the participants and the nonparticipants from the 65 high MAS scorers were 25.92 and 26.78, respectively. This difference was not statistically significant. The MAS scores of the nonanxious group ranged from 0 to 4. The means for the participants and the nonparticipants from the 65 low MAS scorers were 2.46 and 2.31, respectively. This difference was not statistically significant.

Significant differences were found between high and low MAS scorers for three of the manifest needs in men and for five of the manifest needs in women. For men, the mean EPPS scores for the low and high groups, respectively were

8.74 and 11.78 for succorance ($p < .05$), 18.58 and 14.35 for dominance ($p < .01$), and 8.47 and 12.26 for abasement ($p < .01$). For women, similar comparisons were 13.55 and 10.87 for deference ($p < .05$), 19.20 and 16.13 for intraception ($p < .05$), 9.90 and 15.47 for abasement ($p < .01$), 12.85 and 8.20 for endurance ($p < .05$), and 10.25 and 13.27 for aggression ($p < .05$).

Abasement is the only manifest need for which significant differences exist for both men and women in the present study. The finding that both men and women nonanxious subjects scored significantly lower than anxious subjects on abasement suggests the possibility that the non-anxious groups may be characterized by a more favorable self-concept than the anxious groups.

The data for five of the six remaining EPPS variables on which significant differences were found were compared with Edwards' normative data. Edwards has reported significant differences between men and women in his normative data for all five of the variables compared. These comparisons showed that high MAS scorers differed from low MAS scorers on the variables of succorance and dominance for men and deference, intraception, and aggression for women in that the means of the anxious groups were more similar to the scores made by the opposite sex than to those made by the same sex in Edwards' normative group.

This result and the finding of lower scores on abasement for the nonanxious subjects appear to support the view that, as compared to low scorers, high scorers on the MAS are persons who are not as interested in making a good impression on others and may thus tend to say deviant things about themselves.

(Received September 28, 1959)

¹ This investigation was supported in part by a grant from the Penrose Fund of the American Philosophical Society. An extended report of this study may be obtained without charge from Carl D. Williams, Department of Psychology, University of Miami, Coral Gables, Florida, or for a fee from the American Documentation Institute. Order Document No. 6290 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

THE INFLUENCE OF ORDER OF ADMINISTRATION ON SELF-CONCEPT MEASURES¹

DAVID S. PALERMO

Institute of Child Development and Welfare, University of Minnesota

AND JOHN G. MARTIRE

Southern Illinois University

Many studies have reported relationships between measures of actual- and ideal-self-concept and have frequently related measures of self-concept to other indices. This study is designed to investigate the effects of the order of administration of actual- and ideal-self-concept measures on their relationship to each other and their relationship to a third measure.

Self-concept measures were administered to 120 male and 120 female introductory psychology students at Southern Illinois University. Each S was given three trait ranking forms and a set of anchoring definitions for 26 traits used in previous studies by the second author. The Ss were read the anchoring definitions and then instructed to read the directions at the top of the first sheet and make their rankings accordingly. When they finished the first set they were to go on to the second and then the third. Differential instructions at the top of the three ranking sheets directed Ss to rank the traits so as to provide an assessment of self-ideal (sI), actual-self (As) and social-self (Ss). Half of the female and half of the male Ss responded to the rankings in the order As, sI, and Ss (Order I), and the other half of the Ss responded in the order sI, As, Ss (Order II). These orders were selected since most current interest is in ideal- and actual-self concepts.

Pearson r 's obtained between the two orders by sex group for the mean rankings of the 26 traits of the three scales ranged between .82 and .91 (all significant $< .01$ level). These corre-

lations indicate that the relative position of each trait tended to remain the same on all three scales for both sexes regardless of order of administration.

Interscale correlations for each sex group under each order ranged between .26 and .88. These correlations reached the .05 level of confidence with but one exception. The male correlation under Order I between As and Ss (.26) was not significant. Of particular interest are the correlations between Ss and the other two measures under the two orders. Ss correlated .70 and .88 for males and females, respectively with the sI scale when the sI scale was given second, but only .58 and .74 when the sI scale was given first. The correlations between Ss and As were .52 and .68 for males and females, respectively when As was given second and .26 and .62 when As was given first. In each of the comparisons the correlation between the Ss scale, given third in all cases, and the other scales varied as a function of the order of administration. Although the difference between the two correlations was not significant in any of the four comparisons, the consistency of the relationship would suggest that an intervening scale tends to reduce the relationships obtained. If this trend is substantiated by additional research, doubt may be cast upon the validity of discrepancy scores as conventionally computed.

The correlations between the sexes on the rankings of the traits for the three scales ranged between .81 and .89 (all significant $< .01$ level). These data indicate that males and females show a great deal of similarity in their rankings of the traits on all three scales regardless of order of administration. The correlations obtained between sexes in Order II are strikingly similar to those previously obtained by Martire and Hornberger, although the interscale correlations for the two sexes were somewhat different than those obtained in the earlier study.

(Received October 5, 1959)

¹ An extended report of this study may be obtained without charge from David S. Palermo, Institute of Child Development and Welfare, University of Minnesota, Minneapolis 14, Minnesota, or for a fee from the American Documentation Institute. Order Document No. 6292 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

WECHSLER-BELLEVUE PATTERNS OF PSYCHOPATHS¹

LEONARD KINGSLEY

*Walter Reed Army Hospital, Washington, D. C.*²

Wechsler (1944) reported that psychopaths generally obtain higher IQs on the Performance (P) Scale of his test (W-B) than on the Verbal (V) Scale and also described a pattern of subtest characteristics for psychopaths. There have been several W-B studies of psychopaths and delinquents with differing findings. Wechsler (1958) again offers the same IQ and subtest characteristics, qualifying his observations to adolescent sociopaths, but then adding "further experience has shown that it is also applicable to the adult male psychopath."

Twenty psychopathic prisoners in a military disciplinary barracks were matched individually for age and education with a group of prisoners considered to possess "no psychiatric disease." Both groups were later combined into a general prisoner category because of the similarity of the results and matched individually for education and intelligence with "normal" controls functioning in a military setting.³

P IQ was greater than V IQ, at 1% level of confidence, in all 3 groups. This difference was not significantly greater in the psychopaths as compared to the other prisoners nor in the general prisoner group as compared with the controls.

The pattern of subtests postulated by Wechsler was then considered to represent a theoretical

rank order of subtest. The Average Spearman rank order correlation between the individual or combined Wechsler scores and the theoretical rank order was then computed for all 3 groups; the correlation was significant for all 3. The difference scores between each matched subject of each group for the subtests were then computed to see if this correlation was significantly greater in one group vs. the other; these differences were not significant.

Finally, the attempt was made to see if any of the individual W-B subtests, disregarding rank order, differentiated one group from the other; none of the subtests was significantly different in the psychopath vs. prisoner groups or in the general prisoner vs. control groups.

The results of this study raise doubts about the use of difference scores and subtest patterns on the W-B for psychopaths. Wechsler's observations as to the difference between the P and V IQs and the order of subtest difficulty were found to be present for the psychopaths, but these differences and patterns were also present for the nonpsychopathic prisoners and the "normal" controls. The same general trend existed in all three groups, with no significant differences. It is concluded that the IQ differences and subtest patterns offered by Wechsler and based upon the results of adolescent psychopaths cannot be extended to adult psychopaths in a military setting.

REFERENCES

- WECHSLER, D. *The measurement of adult intelligence*. (3rd ed.) Baltimore: Williams & Wilkins, 1944.
- WECHSLER, D. *The measurement and appraisal of adult intelligence*. (4th ed.) Baltimore: Williams & Wilkins, 1958.

(Received October 15, 1959)

¹ An extended report of this study may be obtained without charge from the Secretary, Clinical Psychology Service, VA Hospital, Brooklyn 9, N. Y. or for a fee from the American Documentation Institute. Order Document No. 6294 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

² Now at VA Hospital, Brooklyn, N. Y.

³ The author is indebted to Ernest K. Montague for permission to use these W-B protocols.

THE FACTORIAL VALIDITY OF ITEMS ON THE IPAT ANXIETY SCALE¹

A. W. BENDIG

University of Pittsburgh

The IPAT Anxiety Scale (Cattell, 1957) is composed of 40 items divided into five subscales with each subscale assumedly measuring one of Cattell's five oblique first-order factors (Q3, C, L, O, and Q4) whose intercorrelations define Cattell's second-order factor of "anxiety" (UI 24). These items were assembled into this new scale on the basis of their factor loadings in other inventories, and an analysis of the factor structure of the items in this form has not been reported to date.

The scale was anonymously administered to 200 undergraduate and graduate students in educational psychology classes (100 men and 100 women, mean age 27.8 years), and the items were intercorrelated by the usual product-moment method. Five factors were extracted by the complete centroid method, and the factors were rotated to oblique simple structure using the analytic criterion of rotation developed by Pizka and Saunders. The correlations among the oblique first order factors were factored and the second-order factors also rotated to oblique simple structure.²

¹ An extended report of this study may be obtained without charge from A. W. Bendig, Department of Psychology, University of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. Order Document No. 6295 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

² The author's appreciation is extended to Gary Lotto and William B. Kehl of the University's Computation and Data Processing Center for providing facilities for the statistical analysis.

The obtained item factor loadings showed little relationship to the assumed factor content, and none of Cattell's five factors could be clearly identified in this analysis. Of the 40 items 16 had no loadings above .29 on any of the factors. The correlations among the primary factors ranged from .33 to -.12, and two second-order factors were found instead of the unitary UI 24 factor expected with the correlation between these two oblique factors being approximately .01. Bifactor graphs indicated that further hand rotation would not materially improve simple structure or increase similarity to Cattell's assumed factor structure.

Internal consistency reliabilities were obtained for four of the scores suggested in the test manual (Cattell, 1957): Total Anxiety, Covert Anxiety, Overt Anxiety, and the Difference Score (Bendig, 1959). Using Kuder-Richard Formula 20 the reliabilities of these scores for this sample of 200 Ss were .81, .63, .76, and .24.

The results, unless attributable to a sampling or methodological artifact, strongly suggest that the IPAT Anxiety Scale does not measure what it purports to measure, namely the unitary UI 24 second-order factor of anxiety, but instead confounds two independent second-order factors.

REFERENCES

- BENDIG, A. W. College norms for and concurrent validity of Cattell's IPAT Anxiety Scale. *Psychol. Newsltr, NYU*, 1959, 10, 263-267.
 CATTELL, R. B. *Handbook for the IPAT Anxiety Scale*. Champaign, Ill.: Institute for Personality and Ability Testing, 1957.

(Received October 16, 1959)

ERRATUM

In our article, "Intra Q Deck Relationships as Influences and Realities in Personality Assessment," this JOURNAL, 1960, 24, 61-66, Table 3, column 2, line 1 should be I** instead of II**. This error represented a reversal of the correct interpretation of the relationship between sizes of two correlations. In addition, the third sentence under the subheading, "Neurotic Stereotype and Personality test findings," p. 63, contains an interpretation of this reversal and should read . . . "was described as less neurotic . . ." rather than "more." The remaining materials in the paragraph should then be discounted. Robert D. Wirt and Peter F. Briggs.

OBJECTIVE CONFIGURAL RULES FOR DISCRIMINATING PSYCHOTIC FROM NEUROTIC MMPI PROFILES¹

PAUL E. MEEHL

University of Minnesota

AND

W. GRANT DAHLSTROM

University of North Carolina

One contribution which psychologists are expected to make toward the clinical assessment of psychiatric patients is helping determine the presence or degree of psychotic tendencies. This problem presents itself for reasons of varying pragmatic import, such as: What formal (nosological) diagnosis should be given to the patient? To what extent is there a danger of serious acting out if this patient is seen on an outpatient basis? Is there a psychotic process or structure behind the superficially neurotic manifestations, such that methods of treatment appropriate for neurotic patients are likely to be inefficacious or even deleterious? What is the long-term prognosis, requiring consideration in vocational and educational guidance, advice to relatives, recommendation to a rating board, communications to a family physician, social agency, or court? Should a conservative staff feel entitled to utilize the more radical kinds of treatment, e.g., regressive electroshock, when their policy is to avoid such procedures in the treatment of psychoneuroses? Is this the sort of patient who should probably be treated by Dr. X, who is especially gifted in the treatment of mild schizophrenic conditions and, unlike our other two available therapists, has a preference for working with them? We shall not discuss the utility in clinical decision-making of answering these questions. Suffice it to say that the assessment of "psychotic tendency," phrased in one form or another, is one of the tasks with which the clinical psychologist in most installations will at times be confronted.

The MMPI is among the tests utilized for this purpose. If patients could be effectively

sorted into nosological categories simply by identifying their highest MMPI score, each of the rubrics in the psychiatric nomenclature having a one-to-one correspondence with the MMPI variables as named, the present investigation would be pointless. It is, however, well known that this simple procedure does not work. This is one reason why clinicians prefer to characterize profiles by code rather than by the original scale names. In the University of Minnesota Hospitals the mimeo'd profile form used in patients' charts has *only* code designation for scales. (See Hathaway, 1947; Hathaway & Meehl, 1951a, 1951b; Meehl, 1950b; Welsh & Dahlstrom, 1956.) It is true that one can do significantly better than chance by paying attention only to the highest one or two *T* scores (Hathaway & Meehl, 1951a; Meehl, 1959a) but the amount of improvement over chance, while it testifies to the presence of some validity in the instrument, is not great enough to be very useful. Clinicians even moderately familiar with the MMPI have, therefore, been accustomed for over a decade to interpreting the results by paying attention to the *profile pattern* in some kind of joint relation to the overall elevation and especially the elevation of the most deviant scores.

While it would not be surprising to find that the "clinical eye" had trained itself to recognize configurations not readily identified by conventional linear methods of statistical analysis (Horst, 1954; Lubin & Osburn, 1957), it might be presumed that the clinician's subjective judgment however experienced, assigns less than optimal weights. In addition to this systematic bias, the human judge inevitably throws in some more or less random error variance due to his unreliability. In trusting the clinical eye (at that stage

¹ This study was supported by a Grant-in-Aid of Research from the University of Minnesota Graduate School.

of the total decision-making process which is concerned to classify the profile, although not necessarily the patient!) we treat the nonoptimality of the clinician's un verbalized function, and the temporal instability in its application, as a price paid in order to get the advantages of a configural approach to the profile pattern. Configurality tends to make the clinician's global judgment effective; nonoptimal weights and temporal fluctuation tend to make it ineffective. The net efficiency of the clinician's "judgmental" classification or ordering of profiles is the outcome of these conflicting forces (Meehl, 1959a). For this reason, the present investigation has, in addition to its technological aim of providing an aid to MMPI users, an intrinsic methodological interest. To what extent can the complex configural generalizations exhibited by the behavior of MMPI-skilled clinicians be frozen into a set of clerical operations? If this could be done, it would reduce the amount of clinical experience required by the test user; more importantly, it is probable on theoretical grounds that if the essential features being reacted to by the clinician can be dealt with in an actuarial way, even the asymptote of correct decisions will be increased (Estes, 1957, p. 615).

More important in the long run than either the immediately pragmatic or methodological interests, however, is the possible construct validity (APA Test Standards Committee, 1954; Cronbach & Meehl, 1955) of formal criteria for the "psychotic profile." Psychometric devices should ultimately reach a point of development comparable to the laboratory techniques of internal medicine (biochemical tests, biopsy, roentgenology) such that they are thought of technologically as on the same level with the clinical interview, or ward ratings, rather than as clever devices for predicting these latter "criteria" (Meehl, 1959b). When a patient responds to the verbal stimuli which constitute the MMPI pool in a manner characteristic of previously studied patients recognized by the familiar clinical criteria as schizophrenic, but the clinical staff are not inclined to so diagnose him because of the absence of the traditionally emphasized—and, from Bleuler's point of view, mainly second-

ary—"frankly schizophrenic" symptoms, it is debatable which of these facts about the patient's behavior should be given the greater weight as a probabilistic indicator of his internal psychological state, structure, and dispositions. The development of an objective set of profile pattern criteria for the identification of *recognized, diagnosed* manic-depressive and schizophrenic psychotics has numerous possibilities with regard to subsequent lines of investigation directed at achieving a "bootstrap effect" (Cronbach & Meehl, 1955, p. 286). We inject this methodological note to provide adequate motivation for the contaminated derivation procedures employed.

In what follows, the terms "hit" and "miss" will be employed in the sense of concurrent validity, except where specifically mentioned otherwise. Hints as to construct validity appear in certain of the cross-validation samples, but our main purpose is to present the available *concurrent* validity data for the reader to interpret and utilize however he sees fit within his own pragmatic and theoretical framework. To avoid the tedious repetition of quotation marks, the words "miss" and "hit" will hereafter be employed without them.

DERIVATION OF THE CONFIGURAL RULES

The ultimate justification for a searching procedure is the cross-validated success of the final product. We will indicate only briefly the general working assumptions and provisional hypotheses which underlay our derivation methods. A guiding assumption which would not be admitted by all workers in psychopathology was that there exists some degree of objective typology or taxonomy in a psychiatric population, which will be reflected in the occurrence of profile groups exhibiting a tendency to a kind of "psychometric discontinuity." We did not anticipate that a continuous function of the MMPI variables could be invented which would be of manageable complexity and stable parameters and yet do justice to the configural effects. For example, it seemed doubtful that the weights optimal for making the psychotic-neurotic discrimination *within* profiles exhibiting a 27

code (commonly found both in psychotic depressions and in neurotic depressive reactions or anxiety states) would be very close to the weights optimal for distinguishing between psychotic and neurotic patients exhibiting less self-concern and subjective discomfort and who handle their anxiety via somatizing or projecting mechanisms and present profiles peaked at 3, 4, or 6. So a mixed summative and successive-hurdle model was used throughout in preference to a pure summative model, continuous variables and difference scores being distributed separately for analysis *within* relatively more homogeneous groups initially set apart on the basis of crude but configural criteria such as the Hathaway code. We further assume that because of defects in current nosological concepts (which must, of course, be distinguished methodologically from unreliability in the clinical application of intrinsically powerful diagnostic concepts), or failure of the MMPI scores to provide discriminating information, profile configurations occur which are genuinely without differential significance. Therefore three classification outcomes were allowed: A curve would, by the application of the rules, be classified as either "Psychotic," "Neurotic," or "Indeterminate" in form. A third working assumption was that the formal diagnosis, while it is a readily available crude criterion by means of which patterns are initially identified, is far from infallible, so that we are justified in assuming that sometimes the test is right and the official diagnosis is wrong. Consequently, one must avoid the temptation to formulate too many ad hoc rules in the effort to capture every test miss. With this in mind the non-psychometric case data were deliberately allowed to influence our MMPI rule-making at all stages prior to final cross-validation. Attempts to capture a miss by devising (or revising) a rule were not pursued in those instances in which critical reading of the case material justified serious doubt as to whether the criterion in this case was performing better than the MMPI. The skilled clinical eye was employed as a searcher and idea-originator; statistical runs were employed both as searchers and as checks upon the deliverances of the clinical eye. A fourth working assumption

was that only the most obvious test-related theoretical or dynamic assumptions should play any important part in accepting a rule. We take it for granted that present knowledge, either of personality dynamics or test-taking behavior, is rarely sufficient to justify the neglect of a statistical finding; so that if a certain pattern "works" it is the task of theory, now or in the future, to explain why it does.

On the basis of a preliminary run on the MMPI profiles of 41 male neurotics and 39 male psychotics who had been hospitalized in the inpatient service of the University of Minnesota Hospitals prior to 1946, we chose two sets of difference scores. *Pt* and *Hs* were paired, both having originally been derived as "neurotic" scales but it being part of MMPI lore and general psychiatric experience that the obsessional veers more toward the psychotic side than does the patient with a preference for somatic symptoms. *Sc* and *D* were also paired as a result of this preliminary study. Both of these scales were originally derived on psychotic patients but it is known that *D*, a "mood" scale, is often elevated in neuroses and is markedly so in many cases diagnosed anxiety-neurosis. Severe anxiety states run a very high *D*, yet some of these cases are (as shown by subsequent course or atypical symptoms) actually schizophrenic. It seemed plausible that one might counter-balance these factors by pairing the *D* scale with the more psychotic *Sc* scale. A preliminary cross-validation on 99 additional cases (55 neurotic and 54 psychotic) resulted in approximately 67% hits, so these two difference scores were retained.

We next plotted (*Sc* - *D*) against (*Pt* - *Hs*) on an expanded sample of 104 psychotics and 128 neurotics; five diagonal "bands" were set up on this plot by drawing lines at 45° so located that the central band contained approximately equal numbers of psychotics and neurotics, the extreme bands a distinct preponderance of one group (75% psychotics in the upper right corner and 83% neurotics in the lower left) and the intermediate bands, two and four, showed a trend but with numerous misclassifications. The equations of these lines, with the bracketing

of the four variables rearranged for easier computation, involve the quantity denoted *Beta* in the final rules, and the five regions on this graph correspond to the five bands coordinated to these *Beta* values. $Beta = (Pt + Sc) - (Hs + D)$.

After a study of the cases missed an inspection of several possible distributions of difference scores selected for nonoverlapping with the first two, two other pairs were chosen. $(Pa - Hy)$ showed a good separation among a subset of cases not discriminable by the band method, and had several armchair features to recommend it. Neither of the component scales contains a suppressor correction (McKinley, Hathaway, & Meehl, 1950; Meehl & Hathaway, 1946); both scales have numerous "subtle" items involving denial of pathology and the self-image of normality, rationality, and psychic health (Gough, 1954; Meehl, 1950a; Meehl & Hathaway, 1946; Seeman, 1952, 1953; Wiener, 1948, 1951); both are extrapunitive or impunitive rather than intropunitive, tending to manipulate the environment or to develop symptoms and traits involving less subjective discomfort and anxiety. Both are associated with an inability to "think psychologically," marked lack of insight, and resentment of psychiatric exploration. The $(Pa - Hy)$ difference score might be thought of as located on a psychotic-neurotic axis different from that represented by $(Pt - Hs)$.

The second new difference score was $(Pd - Hs)$. Here an indicator of the acting out, extrapunitive, "tough" component (as contrasted with the more decompensated, suffering and self-concerned individual given to elevations on *Pt* and *Sc*) is balanced off against a somatizing and often passive-aggressive neurotic element. These two scores undergo an approximately equal *K* correction. The sum of the two new difference scores (regrouped for computational ease as the difference of two sums) is designated "Delta" in the final rules. $Delta = (Pd + Pa) - (Hs + Hy)$. We tried thus to counterbalance several difference scores so that the various forms of predominantly psychotic or neurotic mode of adaptation could have a chance to be exhibited in suitable contrast effects. These admittedly loose but clinically plausible consid-

erations were, of course, in each case checked against empirical distributions.

A criterion sample of 262 male patients, evenly divided between neurotics and psychotics, constituted the main basis for rule development. One hundred eighty-seven of these profiles came from the inpatient files of the University of Minnesota Psychopathic Unit; 49 cases were drawn from records available at the VA Hospital at Fort Snelling, Minnesota; and 26 cases were drawn from the files of a Canadian mental hospital. Records with $? \geq 60$, $L \geq 70$, or $F \geq 80$ were excluded. An effort was made to exclude cases in which the MMPI was for some reason (such as catatonic untestability) not taken until the patient had recovered from an acute episode, but it is not always possible to determine this from the records.

Preliminary to the statistical study of these profiles, one of us (PEM) went through the (randomized) set of profiles inspectionally and classified each as neurotic or psychotic. Several months later this procedure was repeated. The results of this sorting will be referred to as the "impressionistic profile classification," and it was part of the evidence used in deciding whether an attempt should be made to write a special rule, or modify a tentative rule, in the effort to eliminate a test miss.

Cases falling within each band (*Beta* region) were distributed as to *Delta* values. While these distributions usually indicated respectable statistical validity, from the standpoint of clinical practice they showed misses felt to be avoidable on the basis of inspection and previous clinical experience. Furthermore, there appeared to be obvious "holes" and deviant "clumps," suggesting the presence of strong minority profile types which might be readily identifiable by features not reflected in *Delta*. Several searching aids were used, such as reshuffling of the profiles so that they would be laid out on the floor in a different random order, or arranging them in different systematic orders from right to left (e.g., first by absolute elevation of the peak score, then by the first digit of the code, then in order of the *Delta* value, then in clusters based upon the formal diagnosis).

Profiles falling wholly within the "normal

range" (i.e., having no $T \geq 70$) behaved differently enough so that a special rule employing the Welsh internalization ratio (Welsh, 1952) was applied to them. A special rule was also invented for mild "fake good" ($L \geq 60$) profiles.

These searching procedures resulted in a set of 13 rules, arranged in sequence so that in applying the system one reads through the list of rules in order until he comes to one which covers the profile under consideration. Application of the rule to this profile then results in one of four outcomes: Profile "psychotic" (P), profile "neurotic" (N), profile indeterminate (I), or "proceed to next relevant rule." In this manner all profiles are finally classifiable under one of the first three rubrics P, N, or I.

This provisional set of rules was then tried out on a preliminary cross-validation sample. Because of the great temptation to capitalize upon sampling errors in such complicated and variable searching procedures as those employed, one expects a marked degree of shrinkage upon cross-validation, and therefore this first cross-validation sample was conceived of only partly as a check on whether we were getting anything but chiefly as a means of modifying the rules by the elimination of ad hoc subrules or adjustment of cutting scores. The preliminary cross-validation sample consisted of 140 file cases of white males aged 18-65 whose MMPI profiles were presumed valid by the above-mentioned *?-L-F* criteria and who were not tested in remission so far as could be judged from the staff notes and other data in the chart. None of these patients had received shock therapy prior to testing, and most were tested within two to three days of admission. Ninety-two of the cases originated from the psychiatric unit of the University of Minnesota Hospitals and 48 from the Minneapolis VA Hospital. All were inpatients. The distribution of diagnoses from the combined criterion and preliminary cross-validation samples ($N = 402$) is shown in Table 1.

Prior to application of the rules, one of us (PEM) read in random order the case summaries of the 92 cases from the university hospital files, deleting material on the psychological test data, the sections presenting dif-

TABLE 1
DISTRIBUTION OF DIAGNOSES ON 402
DERIVATION CASES
(Original criterion plus preliminary
cross-validation)

Neurotics ($N = 201$)		
	<i>N</i>	%
Hypochondriasis	43	21.4
PN mixed	37	18.4
Anxiety neurosis	38	18.9
Reactive depression	31	15.4
Hysteria	29	14.4
Obsessive-compulsive	16	8.0
Other	5	2.5
Neurasthenia	2	1.0
	201	100.0
Psychotics ($N = 201$)		
	<i>N</i>	%
Schizophrenia	107	53.3
Paranoid	(48)	23.9
Simple	(23)	11.4
Hebephrenic	(8)	4.0
Catatonic	(9)	4.5
Mixed	(4)	2.0
Other	(15)	7.5
Paranoid State	22	10.9
Paranoia	2	1.0
Manic-depressive manic	19	9.4
Manic-depressive depressed	28	13.9
Involutional psychosis	23	11.5
	201	100.0

ferential diagnostic considerations, and the final official diagnosis and prognosis. The reader made his own diagnosis on the basis of this reading. The purpose of this procedure was to provide information on criterion trustworthiness in cases where one of the preliminary rules yielded a test miss and a decision had to be made as to whether this particular miss was real or apparent. If the case reader agreed unreservedly with the psychiatric staff and the latter's diagnostic summary did not raise any doubts as to the diagnosis, a more persistent effort was made to modify the preliminary rule so as to avoid the miss than if he disagreed with the staff. We were particularly concerned to detect those cases in which the case reader diag-

nosed psychosis in opposition to a psychoneurotic label attached by the staff, but in which the staff diagnostic summary also included mention of the likelihood of such subclinical diagnoses as, e.g., "incipient schizophrenia" or "strong cyclothymic element."

In the preliminary cross-validation sample the MMPI profiles were again impressionistically sorted as "psychotic" or "neurotic" on two occasions (separated by several months' time and randomizing their order), and the consistency or disagreement of the two sortings was utilized in study of the individual test misses.

As expected, the shrinkage on cross-validation was pronounced. In the criterion sample, 13% of cases were classified indeterminate, and of the remaining 87% of cases for whom a decision was provided by the preliminary rules, the hit-rate was 89% (for both the neurotic and psychotic subpopulations). On preliminary cross-validation, the proportion of indeterminate profiles does not change significantly (10%), but the hit-rate falls dramatically. The hit-rate for neurotics is only 47%, for psychotics 63%. The overall hit-rate is 55%, yielding a hit-rate of only 61% among the 126 cases for which a determinate classification was made. We therefore see a decline of almost 30% in the hit-rate among determinate cases in moving to this new sample. The cross-valid hit-rate, while statistically significant, is pragmatically unimpressive.

Modification of rules, addition of three new rules, and slight alterations in rule sequence were made on the basis of these results. Again the procedures were complex and variable, and concentrated attention upon those rules in which the hit-rate was poorest. Seven of the original 13 rules were in this manner subjected to some kind of modification, ranging from slight adjustment of a cutting score to fairly radical revision. Constant back-reference was made during this phase to the effect of proposed modifications upon the hit-rate in the original 262 criterion cases.

One interesting result of these procedures was the progressive displacement of the originally emphasized band rules to the end of the rule sequence. The patterns of six scales represented by Beta and Delta only begin to

function as powerful discriminators after the curves have initially been divided into major "types." One is reminded here of the point made by Block (1957) in another context, that the kind of *R* covariation which one is likely to discover by traditional individual-differences methods may be quite misleading unless the population has first been divided into subpopulations by *Q* covariation analysis, the *R* covariation patterns being sometimes very different within the several subpopulation "types."

The modifications made resulted in a drop in the criterion group from 87% to 77% determinate classifications with a maintenance of 89% hits among cases classified; in the cross-validation group there was a decline from 90% to 74% determinate classifications, but an associated increase in the hit-rate from 61% to 83% among those classified. For the total augmented criterion group ($N = 402$), we have rules which leave 24% of the cases indeterminate but give us an 87% confidence when a classification is made. Thus the movement in modification was in the direction of increased recognition of nondifferentiating patterns, liquidating ad hoc rules which served to classify special criterion cases but which collapsed on cross-validation. The final set of rules is presented in Appendix A.²

CROSS-VALIDATION

For final cross-validation of the modified rules we attempted to secure a sample which would be both large and sufficiently diverse (as to clinical populations and staff diagnostic practices) to provide some information about validity generalization. Through the kindness of several clinical psychologists in providing MMPI data we were able to obtain eight different samples of cross-validation cases from

² A five-page statement of the rules has been deposited with the American Documentation Institute. Order Document No. 6330 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress. Mimeographed copies of the rules may be obtained from Paul E. Meehl, Box 390 Medical School, University of Minnesota, Minneapolis 14, Minnesota.

various clinical installations over the country.³ The sample sizes varied from 42 to 273, with a median of 97 cases per sample and a total $N = 988$. A more detailed characterization of these samples is contained in Appendix B.⁴ The samples vary as to nature of the population (VA and non-VA, outpatient and inpatient). They are geographically dispersed, and so far as known to us there is considerable variation as to the predominant local theoretical and diagnostic orientation. *All, however, are males.* They vary widely in the diagnostic "purity" of the cases. In three samples (A, B, E) the criterion is completely uncontaminated, the MMPI having been unavailable to the diagnosing clinicians. A fourth (K) is effectively uncontaminated since, although MMPI results were available, the patient population sampled consisted of psychotics in a state hospital where the daily census of neurotic diagnoses runs only 1-2%. The remaining four samples (C, D, F, G) suffer from unknown but nonnegligible contamination. *Hit-rates do not differ as between the contaminated and uncontaminated cases* ($\chi^2 = 1.13$, 1 *df*, $p > .20$). Tables 2 and 3 summarize the results of applying the rules to these eight samples.

The H:M:I distribution does not vary significantly over the seven neurotic categories (omitting the single phobic reaction, $\chi^2 = 20.84$, 14 *df*, $.10 < p < .20$). Pooling all schizophrenics, H:M:I does not vary over psychotic diagnoses ($\chi^2 = 8.08$, $p > .80$). Although inspection suggests that the affective

disorders as a group are harder to identify than the schizophrenics, this trend does not quite achieve statistical significance ($\chi^2 = 5.50$, $.05 < p < .10$). Further checking on larger numbers of involuntaries and manic-depressives is indicated especially since among the *decidable* cases, hit-rates ($H/H + M$) do differ as between schizophrenics and affectives (.75 versus .54, $\chi^2 = 4.97$, 1 *df*, $.02 < p < .05$). That is, when we ignore indeterminate curves, the affective cases are more often misidentified as neurotic. Qualitative study of the missed affective cases suggests that criterion error or patient change accounts for some of these misses.

The variation of H:M:I over schizophrenic subcategories fails to reach statistical significance ($\chi^2 = 14.67$, 10 *df*, $.10 < p < .20$). It is interesting to note that the best results, in terms of high total hits, low indeterminate rate, and high confidence among classifiables, occur in the conversion and somatization groups. As we move into neurotic categories where dysphoria, failure of the defense, intellectualizing and other obsessional mechanisms, conscious guilt, and "bad" self-concepts are more in evidence, the profiles increasingly resemble those of the psychotic group.

The H:M:I distribution does not differ significantly as between all psychotics and all neurotics ($\chi^2 = 2.39$, 2 *df*, $p > .30$). Considering only the 691 determinate-curve cases, neurotics and psychotics do not differ in hit-rate ($\chi^2 = 2.35$, 1 *df*, $.10 < p < .20$).

The variation in H:M:I over the eight samples is significant ($\chi^2 = 43.14$, 14 *df*, $p < .001$). One sample, D (Rubin's Chillicothe data), fails to exceed chance differentiation among classified cases. All samples do show $H > M$, the ratios varying from 1.6 to 15.6. The hit-rates as between neurotic and psychotic cases do not differ in seven of the samples, but the 19% higher identification of neurotics in Sample G is at the .001 level.

Whether one is heartened or discouraged by these results will depend both upon his clinical expectations and his methodological orientation. If we concentrate upon the fact that the total number of correct classifications in the entire sample is only 53.3%, we may not be much impressed. If, however, we accept the fact that some curves are ambiguous

³ We wish to express our indebtedness to the following clinical psychologists who were of great assistance to us through their kind efforts in making their data available to us for reanalysis or for taking the pains to track down diagnoses, original records, and other information on the criterion and cross-validation samples: H. R. Albrecht, VA Hospital, Chillicothe; George Guthrie, Pennsylvania State College; Howard F. Hunt, University of Chicago; Thomas Kiresuk, Minneapolis General Hospital; Timothy F. Leary, Kaiser Foundation, Oakland, California; James C. Lingo, Langley Porter Clinic; Donald R. Peterson, University of Illinois; Albert Rosen, University of Maryland; Harold Rubin, VA Mental Hygiene Clinic, Philadelphia; Randal M. Wolfe, Chillicothe VA Hospital.

⁴ A four-page summary of the sources and characteristics of the eight cross-validation samples has been deposited with the American Documentation Institute (See Fn. 2).

and *ought* not to be given any appreciable weight in making a decision (an attitude toward laboratory tests to which our medical colleagues have become thoroughly accustomed), and therefore attend chiefly to the subset of cases for which the rules pro-

vide a decision, we see that the ratio of hits to misses among decided curves is 3.2 to 1. When a curve is classified on the basis of the rules, the classification has attached to it a confidence of .76. Assuming the MMPI to be in use in a clinic for whatever (multiple) pur-

TABLE 2
COMPARATIVE RESULTS BY DIAGNOSIS IN THE EIGHT CROSS-VALIDATION SAMPLES
($N = 988$)

Diagnosis	N	Neurotics				
		%	H%	M%	I%	$\frac{H}{H + M}$
Psychoneurosis or psychosomatic, unspecified	185	35.0	46.5	17.8	35.7	.72
Anxiety	135	25.5	57.0	17.8	25.2	.76
Conversion or somatization	99	18.7	69.7	8.1	22.2	.90
Depression	65	12.3	52.3	13.8	33.8	.79
Mixed	23	4.3	47.8	13.0	39.1	.79
Other (e.g., passive-aggressive, neurotic character)	9	1.7	77.8	0.0	22.2	1.00
Obsessive-compulsive	8	1.5	37.5	12.5	50.0	.75
Hypochondriasis	4	.8	50.0	25.0	25.0	.67
Phobic Reaction	1	.2	100.0	0.0	0.0	1.00
	529	100.0				
		Psychotics				
Schizophrenia paranoid	174	37.9	59.2	14.9	25.9	.80
Schizophrenia mixed or unspecified	79	17.2	41.8	22.8	35.4	.65
Schizophrenia latent or in remission	59	12.9	54.2	18.6	27.1	.74
Schizophrenia hebephrenic	6	1.3	33.3	0.0	66.7	1.00
Schizophrenia simple	6	1.3	50.0	33.3	16.7	.60
Schizophrenia catatonic	3	.7	33.3	0.0	66.7	1.00
Psychosis unspecified	92	20.1	47.8	17.4	34.8	.73
Manic-depressive depressed	12	2.6	33.3	33.3	33.3	.50
Manic-depressive manic	10	2.2	50.0	30.0	20.0	.62
Manic-depressive mixed	2	.4	50.0	50.0	0.0	.50
Involuntal psychosis	8	1.7	37.5	37.5	25.0	.50
Paranoid condition	6	1.3	66.7	16.7	16.7	.80
Psychosis with psychopathic personality	2	.4	100.0	0.0	0.0	1.00
	459	100.0				

TABLE 3
COMPARATIVE RESULTS IN THE EIGHT CROSS-VALIDATION SAMPLES

Sample	N	Criterion % psychotic	H%	M%	I%	$\frac{H}{H+M}$	p^b
A ^a	92	64	55	16	28	.77	<.001
B ^a	77	52	45	29	26	.61	<.05
C	103	47	49	16	35	.75	<.001
D	42	52	40	21	38	.65	ns
E ^a	181	50	45	18	36	.71	<.001
F	166	43	47	20	33	.70	<.001
G	273	27	63	12	25	.84	<.001
K ^a	54	100	78	5	17	.93	No test
Total	988	54	53	17	30	.76	<.001

^a Essentially uncontaminated samples.

^b Chi square or exact test used as appropriate.

poses it is deemed appropriate, the minute or less of clerical time required to apply the Meehl-Dahlstrom rules is not an unjustifiable expenditure of effort to obtain this much additional information on a sizable subset of cases. That the concurrent validity compares favorably with the pooled weighted judgments of 29 Minnesota clinicians (and is better than any of them taken individually) would seem to justify substitution of the rules for impressionistic profile assessment with respect to the psychoticism variable. These patterns also discriminate better than any of five other "actuarial" methods, including the linear discriminant function (Meehl, 1959a).

One of the present samples (Palo Alto, Sample B) does provide sufficient uncontaminated information to make feasible certain *within*-diagnosis comparisons which shed a little light upon construct validity. Through the kind cooperation of Howard Hunt, we were able to obtain access to a one-page psychological summary sheet for each of the patients in that sample. This summary, in addition to the usual face-sheet data, also included varying amounts of information regarding diagnostic considerations (e.g., impression of the admission board, number of previous admissions, type of service discharge, compensation, hospital status at the time of testing, Rorschach or Wechsler-Bellevue findings). One of us (PEM) read over each of these summary sheets (uncontaminated by knowledge of the associated MMPI profile) and made a subjective judgment in six steps as

to the strength or clarity of evidence for and against psychotic tendencies. To get a rating of ++ for "very clear psychosis," the data had to include the admission board impression and a Rorschach diagnosis, neither of which considered any alternative. Semiobjective rules were set up for lesser degrees of clarity or amount of information running through +, ±, = ("neurotic but debatable"), —, and — — ("very clear neurosis, considerable data given"). Any case marked "in remission" was automatically considered questionable. Cases with disharmony between admission board impression and final diagnosis were automatically considered questionable. For the 44 cases judged as either "very clear" or "clear," the hit-rate was 68% (67% and 69% for the two levels of clarity, respectively); whereas for the 13 cases considered "doubtful," only four were correctly classified, corresponding to a hit-rate of 31%. ($\chi^2 = 5.84$, $p < .02$). That is, when the admission board was impressed with some behavior not reflected in the formal diagnosis, or the psychiatric staff was moved to record a secondary diagnosis, or the psychologist giving the Rorschach found evidence running counter to the diagnosis administratively assigned, such a patient was more likely to produce an MMPI profile out of harmony with the official diagnosis than was true for patients in whom no such inconsistencies were in evidence.

The relation between construct and concurrent validity in situations of this type is too

complex and involves too much methodological controversy to be developed here (see Meehl, 1959b). From the standpoint of concurrent validity, the Palo Alto findings just described can be interpreted either favorably or unfavorably, depending upon one's pragmatic emphasis. From a traditional viewpoint, the subjective rating based on the face-sheet data may be viewed as a "criterion" (although highly unreliable); the diagnostic reversals given by the pattern rules in this intermediate range must then be conceived as validly reflecting the "mixed" character of the patient's behavior, certain aspects of which unduly influence the psychiatric staff responsible for the formal diagnosis while other aspects are conveyed by the face-sheet information and therefore reflected in the "clarity" ratings of the case reader. From another point of view, one can argue clinically that the practical utility of a psychometric instrument varies *inversely* as the clarity with which the patient's diagnosis can be pegged without the instrument, so that the finding of high agreement for clear-cut cases and poor agreement for borderline cases is precisely the reverse of what might be desired in clinical practice. The rationale of this second position involves an implicit commitment to the idea of construct validity and brings us to our own preferred mode of thinking about such internal relationships. The construct validity of the profile is its power to reveal the internal psychological structure and state of the patient, which it does fallibly and probabilistically, as do the other indicators available (including the social impact of the patient upon the diagnosing psychiatrist). *To say that one wants the test precisely for cases where the diagnosis is otherwise difficult, is to say that "mixed" behavior output leaves one in doubt as to the inner psychological condition, and it is desired to use the test as an aid in assessing this inferred internal state of affairs.* This way of conceptualizing the situation makes it unfeasible to refine the validation procedure further unless the quantitative and qualitative character of the additional data (other than formal diagnosis) is superior to that routinely available.

Since the kind and degree of construct validity remain to be established and the pres-

ent investigation deals only with the most obvious among the indefinite family of concurrent and predictive validities which the patterns may possess, a vexing problem of terminology obtrudes itself. How shall the decidable profiles be labeled? In the present state of the evidence, we would favor a rather noncommittal language. The adjectives "psychotic" and "neurotic" might be carefully defined as referring merely to *curve types*, warning against an automatic classification of the *patient*, but one realizes that such hygienic semantic provisions do not always achieve their aim in practical usage. Furthermore, the term "psychotic" has itself a rich network of associations, valuable for theoretical purposes and in the design of subsequent research but which in daily clinical practice should not be linked too closely to the curve form itself. We propose as a terminological convention that the neurotic curve *type* as identified by these rules be called an "N-curve," and the psychotic curve *type* be referred to as a "P-curve." The construct, concurrent, and predictive validities associated with these types and the subtypes (rules) by which the broader groups are identified then remain to be determined. This approach is consistent, at the *profile* level, with our preference for the use of digits in designating individual MMPI keys. An alternative and theoretically neutral location would be "first zone" (=N) and "third zone" (=P), reserving "second zone" for the conduct disorder types, adopting a "zone" language as is done, e.g., in the colloidal gold test of clinical neurology.

We hope that a "bootstraps effect" has been achieved by the identification of these configurations. From a construct validity viewpoint, the phenomenon of a "test miss" can be very instructive and in the long run may be more productive of understanding and improvement in assessment procedures than an obvious "hit." Suppose that a patient presents, on formal mental status examination and other behavior data gathered from informants and ward personnel, a nonpsychotic clinical picture. Yet, as happens with considerable frequency, the MMPI pattern is a schizophrenic P-type. Assume exclusion of obviously invalid or essentially "chance" response patterns readily detectable by present

methods. What are we to think about this patient? Setting aside questions of evaluating the test, one can contemplate this situation in terms of the patient's psychology. The patient, when presented in a standard manner with 550 verbal stimuli, has responded to them systematically, in a way which is "statistically congruent" with the response patterns of patients who were *not* clinically clear of psychotic behavior but rather identified by nontest criteria as diagnosable schizophrenics. This systematic pattern of responding cannot be dismissed. In any adequate account of the psychometric situation, it will have to be fitted into *some* kind of consistent causal analysis. When out-and-out faking or lying has been excluded, the verbal responses "true" and "false" to this standardized set of verbal stimuli must be psychologically construed as phenomenological reports of varying adequacy. It is of course not assumed that the reports, insofar as they have a *content* referring to behavioral dispositions or the facts of the world or of other people's conduct, are "correct" (Meehl, 1945). What is, however, assumed—once we have excluded clear fake records—is that the patient is reporting, within the limitations of a prespecified domain and fixed-response context, the character of his phenomenology at the time of testing. We then have a paradox. Here is a patient whose MMPI-phenomenology is "statistically characteristic" of clinically obvious schizophrenia but whose behavior in a diagnostic interview, on the ward, and in recent extramural contexts is not clinically schizophrenic. Such a patient *must* be psychologically different from both the clinically schizophrenic patient with a third zone curve and the neurotic patient with a first zone curve. His verbal response to these items is not something detached from the world and utterly uninterpretable without nontest data. After all, the test has elicited a sizeable mass of verbal behavior, and if previous evidence enables us to say "what is a schizophrenic way of responding to these items," the fact that a patient has responded thus is a weighty piece of evidence regarding him and must be "explained" rather than, as is so often done in staff conferences, "explained away."

There are dangers involved in this kind of

thinking for those who would irresponsibly defend an instrument to which they are strongly committed, effectively cutting themselves off from all possibility of refutation; but this danger should not mislead us into undervaluing the evidential weight of a psychometric pattern. An adequate understanding of the class of patients defined as test-misses from the concurrent validity standpoint must include a psychological account of the presence of their schizophrenic phenomenology, when their social and other behavior is apparently that of the neurotic group. The development of such an account would seem to hinge upon the accumulation of further data on as many different kinds of concurrent, predictive, and content validity as the ingenuity of clinical investigators can devise. Examples which come readily to mind would be the following: In a group of patients all officially diagnosed as psychoneurotic, what is the relationship between the incidence of schizophrenic Rorschach indicators and the presence of P-curves of schizoid type? If we assume that some so-called "neurotic depressive reactions" are actually phases of a "damped" endogenous manic-depressive cycle, are there nontest indicators (e.g., a history of recurring depressions in the patient, or the finding of depressions and suicides in his family history) associated with "neurotic" depressions manifesting P-profiles of the depressive type? What is the relationship between curve type and such physiological indicators as the Funkenstein reaction (Funkenstein, Greenblatt, & Solomon, 1952) or the sedation threshold (Shagass & Jones, 1958)? If we classify patients on the basis of curve type, and then for each curve group, regardless of the clinical diagnosis officially given at the time of initial study, plot cumulative subsequent hospitalization over a long time span, how do the slope and asymptote constants of these curves compare? Suppose that a group of patients seen in an outpatient setting are Q sorted and Q correlations with an idealized Q sort description of "pseudo-neurotic schizophrenia" are computed. Considering a sample of patients all diagnosed as neurotic, how do the distributions of these Q correlations compare as between neurotic patients having a P- versus those having an N-profile? It is by

accumulation of investigations of this type, preferably in the context of at least a sketch of theory as to the nature of the psychotic disorders, that an adequate picture of the construct validity of profile pattern signs will have slowly to be achieved.

SUMMARY

By a combination of statistical searching and codification of clinical experience, a set of objective profile signs has been evolved as an aid in the discrimination between "neurotic" and "psychotic" MMPI profile patterns. These signs were then applied to the profiles of eight cross-validation samples ($N = 988$) from a diversity of clinical installations involving varying degrees of contamination from essentially none for four samples to unspecifiably high. Approximately 30% of the cases present profiles which are classified as indeterminate with regard to the psychotic-neurotic distinction. Among classifiable cases, the concurrent validity hit-rate varied from a low of 61% to a high of 93% in the eight samples, with a median of 73% hits, and a hit:miss ratio of 3.2:1 for the total pooled sample. It is suggested that research with these patterns be directed to the elaboration of their construct validity and to the psychological understanding of the phenomenology and dynamics of cases which are "test-misses" from the concurrent validity standpoint. It is also suggested that curves characteristic of those found in diagnosed psychoneurotics be designated simply as N-type (or "first zone") profiles, in contradistinction to P-type ("third zone") profiles manifesting the configuration typical of psychotic patients. The rules, while complex and of a somewhat forbidding aspect, can, after a little practice, be applied by a clerk in less than a minute's time.

REFERENCES

- AARONSON, B. S., & WELSH, G. S. The MMPI as a diagnostic differentiator: A reply to Rubin. *J. consult. Psychol.*, 1950, 14, 324-326.
- AMERICAN PSYCHOLOGICAL ASSOCIATION, Test Standards Committee. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull.*, 1954, 51, Suppl. No. 2, Part 2, 1-38.
- BLOCK, J. The difference between Q and R. *Psych. Rev.*, 1955, 62, 356-358.
- CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, 52, 281-302.
- ESTES, W. K. Of models and men. *Amer. Psychologist*, 1957, 12, 609-617.
- FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. An autonomic nervous system test of prognostic significance in relation to electroshock treatment. *Psychosom. Med.*, 1952, 14, 347-362.
- GOUGH, H. G. Some common misconceptions about neuroticism. *J. consult. Psychol.*, 1954, 18, 287-292. (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- HATHAWAY, S. R. A coding system for MMPI profiles. *J. consult. Psychol.*, 1947, 11, 334-337. (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- HATHAWAY, S. R., & MCKINLEY, J. C. *The Minnesota Multiphasic Personality Inventory Manual* (Revised). New York: Psychological Corp., 1951.
- HATHAWAY, S. R., & MEEHL, P. E. *An atlas for the clinical use of the MMPI*. Minneapolis: Univer. Minnesota Press, 1951. (a)
- HATHAWAY, S. R., & MEEHL, P. E. The MMPI. *USA Tech. Man. TM 8-242; USAF Man. AFM 160-145*. Washington, D. C.: United States Government Printing Office, 1951. (b)
- HORST, P. Pattern analysis and configural scoring. *J. clin. Psychol.*, 1954, 10, 3-11.
- HUNT, H. F., CARP, A., CASS, W. A., WINDER, C. I., & KANTOR, R. E. A study of the differential diagnostic efficiency of the MMPI. *J. consult. Psychol.*, 1948, 12, 331-336.
- LUBIN, A., & OSBURN, H. G. A theory of pattern analysis for the prediction of a quantitative criterion. *Psychometrika*, 1957, 22, 63-73.
- MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The MMPI: VI. The K scale. *J. consult. Psychol.*, 1950, 14, 165-171. (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- MEEHL, P. E. The dynamics of "structured" personality tests. *J. clin. Psychol.*, 1945, 1, 296-303. (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- MEEHL, P. E. Configural scoring. *J. consult. Psychol.*, 1950, 14, 165-171. (a) (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- MEEHL, P. E. *Using the MMPI in counseling*. Advisement & Guidance Section, VRE. St. Paul: VA, 1950. (b)
- MEEHL, P. E. A comparison of clinicians with five statistical methods for identifying psychotic MMPI profiles. *J. counsel. Psychol.*, 1959, 6, 102-109. (a)

- MEEHL, P. E. Some ruminations on the validation of clinical procedures. *Canad. J. Psychol.*, 1959, 13, 102-128. (b)
- MEEHL, P. E., & HATHAWAY, S. R. The K factor as a suppressor variable in the MMPI. *J. appl. Psychol.*, 1946, 30, 525-564. (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- PETERSON, D. R. The diagnosis of subclinical schizophrenia. *J. consult. Psychol.*, 1954, 18, 198-200. (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- PINTNER, R. Accuracy in scoring group intelligence tests. *J. educ. Psychol.*, 1926, 17, 470-475.
- ROSEN, A. Development of some new MMPI scales for differentiation of psychiatric syndromes. Unpublished doctoral dissertation, Univer. Minnesota, 1952.
- ROSEN, A. Differentiation of diagnostic groups by individual MMPI scales. *J. consult. Psychol.*, 1958, 22, 453-457.
- RUBIN, H. The MMPI as a diagnostic aid in a veterans hospital. *J. consult. Psychol.*, 1948, 12, 251-254.
- RUBIN, H. A note on "Reply to Rubin." *J. consult. Psychol.*, 1950, 14, 327-328.
- SEEMAN, W. Subtlety in structured personality tests. *J. consult. Psychol.*, 1952, 16, 278-283. (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- SEEMAN, W. Concept of "subtlety" in structured psychiatric and personality tests: An experimental approach. *J. abnorm. soc. Psychol.*, 1953, 48, 239-247. (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- SHAGASS, C., & JONES, A. L. A neurophysiological test for psychiatric diagnosis: Results in 750 patients. *Amer. J. Psychiat.*, 1958, 114, 1002-1009.
- WELSH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, 16, 65-72. (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- WELSH, G. S., & DAHLSTROM, W. G. (Eds.) *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.
- WIENER, D. N. Subtle and obvious keys for the MMPI. *J. consult. Psychol.*, 1948, 12, 164-170. (Reprinted: In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.)
- WIENER, D. N. A control factor in social adjustment. *J. abnorm. soc. Psychol.*, 1951, 46, 3-8.

(Received September 16, 1959)

FACTORS IN PSYCHOTHERAPY CHANGE¹

ROBERT C. NICHOLS AND KARL W. BECK

Purdue University

Attempts to determine the effectiveness of psychotherapy have used many different criteria to assess the extent of change and there has been considerable discussion of what variables constitute the most adequate measures. Some researchers have solved the problem of having to decide between a number of available criteria, none of which is completely satisfactory, by combining several measures into a composite criterion. Assumptions implicit in this approach are that psychotherapy brings about a general and unitary personality change and that all criteria are imperfect measures of this change. Other researchers have insisted that the effects of psychotherapy may be very specific and that the criterion for evaluation of therapeutic outcome should bear a direct theoretical relationship to the underlying theory of therapy. It is important to know the degree of generality of therapeutic outcomes, both in making decisions concerning appropriate criteria for research, and in determining whether specific therapies may be needed for specific problems or whether general psychotherapeutic techniques may be applicable to a wide variety of problems.

Two previous factorial studies of therapy change (Cartwright & Roth, 1957; Gibson, Snyder, & Ray, 1955) suggest that personality change is multidimensional and that the factors tend to be organized around the various instruments used to measure change. It is, however, erroneous to interpret this as indicating that the effects of psychotherapy are themselves multidimensional. The therapy is certainly not the only independent variable operating to produce change in a group of subjects (Ss) undergoing therapy. Thus, to study the dimensionality of *therapy* change,

it is necessary not only to determine the factor structure of the change measures, but also to establish which, if any, of the change factors are related to the amount of therapy received.

The ideal design for dealing with this problem would be to obtain a variety of measures of personality change from a large number of Ss undergoing psychotherapy and to obtain the same change measures from a matched group of control Ss over a comparable waiting period. A comparison of the factor structure of the change measures in the therapy group with that in the control group would indicate whether or not therapy has a general or group effect which would tend to increase the correlations among groups of change measures in the therapy group. Then comparisons of the amount of change on the various factors in therapy and control groups would indicate which change factors are affected by therapy and which are not. Unfortunately, this paper is not a report of the results of this ideal study, but rather is an attempt to answer some of the same questions with somewhat less adequate data.

In this study, a number of therapy change scores are factored in order to determine the dimensions along which change occurs in the measures used. Changes along these dimensions in the therapy group are compared with changes in a comparable control group in order to determine which of the changes can be attributed to the therapy. Additional evidence on this point is obtained from the relationship of the changes in the therapy group to the amount and intensity of the therapy.

METHOD

Subjects

During the 1957-58 and 1958-59 academic years, 75 undergraduate students were seen at the Purdue

¹ This study was supported by a grant (XR 1916) from the Purdue Research Foundation.

Psychological Clinic for at least five psychotherapeutic interviews and completed the California Psychological Inventory (CPI) and a sentence completion test before the first and after the last interview. At the end of therapy they also completed a brief rating scale, a portion of which is shown in Table 1, and the therapist filled out a somewhat longer rating scale which included four items corresponding to those shown in Table 1. There were 57 male and 18 female clients. Approximately half were self-referred to the clinic while the other half were referred from the dean's office, the student health service, or other campus agencies. Difficulties presented in the intake interview ranged from alcoholism and homosexuality to lack of motivation to study. The most common complaints were difficulties in interpersonal relationships, difficulties in school work, and family or autonomy problems. None of the cases included in this study were so sick that they had to leave school during the course of therapy. The mean number of therapy interviews was 14.7 and the standard deviation was 9.2.

These Ss were seen for therapy by third- and fourth-year graduate students who were enrolled in a two-semester practicum course. There were 41 therapists, most of whom saw two of the research Ss. Most of the therapists had had some previous experience with psychotherapy, but for a few this was their first direct therapy experience. All of the therapy was closely supervised, typically one hour of supervisory conference with a faculty member for each hour of therapy. Recordings and observation through a one-way screen were occasionally used as an aid to supervision. Most of the clients were seen once weekly, but a few were seen two or three times a week. The goals of therapy varied considerably from therapist to therapist, but for the most part were in the direction of the establishment of a close interpersonal relationship with the client and the discussion of feelings and problems with a view to the client's gaining insight.

When a client came to the clinic he was seen for an initial intake interview by a graduate assistant and an appointment was made to take the CPI and the sentence completion test. After the tests were completed the client was assigned to a therapist who, with his supervisor, was then in complete charge of the case. The clinic closed for the summer at the end of the spring semester, which caused the termination of therapy for 58 of the 75 cases included in the present study. The remaining cases terminated voluntarily some time before the end of the spring semester. Following the last therapy interview the client was contacted by letter or phone by the experimenters (*Es*) and asked to come in to take the posttests. These tests were administered by the *Es* and the clients did not see their therapists at this time.

As might be expected it was often difficult to get all of the required information on every client. Of 157 clients who took the initial tests only 75 met the requirements of the study. Most of those not included were eliminated because of an insufficient

number of interviews. However, a number were not included because they dropped out of school or for some other reason did not take all of the tests.

Factor Analysis of Change Scores

For each of the 75 Ss, 30 change scores were available. The CPI was scored for the 18 standard scales described in the CPI manual (Gough, 1957) and change scores were calculated by subtracting the pretest score from the corresponding posttest score. The sentence completion test consists of 55 stems developed by Rosenberg (1957), each of which presents a conflict situation to be reconciled. The test was scored according to an unpublished system developed by the junior author in which the Ss' reconciliation of the conflict presented by each stem was classified as either active or passive and as either firm (dealing directly with the conflict situation) or soft (avoiding the conflict situation). The combinations of these ratings summed over all items yielded four scores: active-soft (AS), active-firm (AF), passive-soft (PS), and passive-firm (PF). Interscorer reliability was found to be .87 for AS, .75 for AF, .64 for PS, and .49 for PF by Ebner (1958) using a shortened form of the test and another college student sample. Four sentence completion change scores were calculated by subtracting the pretest score from the corresponding posttest score.

Four ratings of amount of change were made at the end of therapy by both therapist and client on the items shown in Table 1. These change scores were intercorrelated and the resulting 30×30 matrix was factor analyzed by the principal components method with communalities estimated as the highest correlation in each row.² On the basis of the size of the largest factor loadings and the size of the latent roots it was decided to retain six factors which were rotated analytically to the varimax criterion of orthogonal simple structure.³ The rotated factor matrix is shown in Table 2.

THE FACTORS

The findings of this analysis are consistent with those of Cartwright and Roth (1957) and Gibson, Snyder, and Ray (1955) that the amount of change occurring with psychotherapy depends to a large extent on the vantage point from which it is observed. Each

² The correlation matrix and the unrotated factor matrix have been deposited with the American Documentation Institute. Order Document No. 6405 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

³ These calculations were done on Purdue's Datatron 205 computer.

TABLE 1
RATING SCALE FILLED OUT BY THE THERAPY GROUP AT THE END OF TREATMENT

We are trying to evaluate the effectiveness of the work of the Psychological Clinic in an attempt to improve the services offered. Your frank answers and comments to the items below will greatly help us in doing this.

On the following four items rate the amount of change, if any, which has occurred in you since you first came to the clinic. Consider all changes whether due to coming to the clinic or not. Indicate your rating by a check mark by the appropriate number.

1. Amount of change in the symptoms or complaints which brought you to the clinic.				
1	2	3	4	5
worse than when I came	no change	some improvement	considerable improvement	no longer bothered by symptoms
2. Amount of change in your understanding of yourself and your behavior.				
1	2	3	4	
more confused about myself now	no change	understand myself some better	understand myself much better now	
3. Amount of change in feeling and general outlook on life. (Happiness, sense of well being, etc.)				
1	2	3	4	5
feel worse	no change	feel some better	feel considerably better	feel a great deal better
4. Overall rating of amount of change considering everything that is important to you.				
1	2	3	4	5
change for the worse	no change	some change for the better	considerable change for the better	a great deal of change for the better

Note.—Therapists and control subjects completed scales very similar to the above. The direction of the scale for Items 2 and 4 was reversed on the actual rating scale so that low scores indicated improvement; however, all scores reported here are in terms of the scales in the above form.

instrument used to assess change defined its own factor or factors.

Factor A is clearly a CPI factor with high loadings on Wb, Re, Sc, To, Gi, and Ac. Since the Wb and Gi scales were constructed to detect faking, this factor may be tentatively interpreted as representing change in the tendency to present oneself in the best light on inventory items. The other scales loading on this factor represent changes in self-description in regard to maturity, conscientiousness, permissiveness, and dependability. However, the profiles presented by Gough (1957) of Ss instructed to fake bad show a marked lowering of the scores on these scales. As a matter of fact, the differences between the *T* scores of Ss faking good and Ss faking bad reported by Gough (1957) are, with the exception of the Cm scale, pro-

portional to the corresponding factor loadings on the present factor.

If the interpretation of this factor as a faking or social desirability factor is correct, it is surprising that the client's direct ratings of change do not load on it. Perhaps this is because the CPI scores are difference scores between pretest and posttest and would thus be sensitive to any change in tendency to present oneself in a good light, while the client ratings are made only at the end of therapy when the client may not be aware of this change in self-description or may not consider it in his ratings.

Factor B is clearly a therapist rating factor. In addition to the therapist ratings, the client rating of change in understanding loads on this factor. What is surprising is that the therapist ratings identify a factor which is so

little related to test score changes and client ratings. The loading of client rating of change in understanding may represent some real agreement between therapist and client. However, since any changes achieved were presumably discussed in the final interviews, one would reasonably expect more agreement than this.

Factor C is another CPI factor on which Do, Sc, Sy, Sp, and Sa have the major loadings. There are lower loadings for Wb, To, Ac, Ai, Ie, and PS. The highest loading CPI

scales on this factor are grouped together by Gough (1957) as measures of poise, ascendancy, and self-assurance. They differ from those loading primarily on Factor A in that they all seem to be concerned with ease of social interaction and comfortableness with others, where the scales loading on Factor A are more concerned with responsibility and self-control.

Factor D is a client rating factor and has generally negligible loadings for all other variables. There is a low loading for the therapist

TABLE 2
ROTATED FACTOR MATRIX

Change Scores	Factors						h ²
	A	B	C	D	E	F	
CPI Change							
Do Dominance	.16	-.10	.57	.03	.24	.34	.53
Cs Capacity for status	.12	.06	.62	.18	-.10	-.08	.45
Sy Sociability	.19	-.13	.71	.05	.25	.18	.65
Sp Social presence	.08	-.03	.75	.19	.14	-.01	.63
Sa Self-acceptance	.03	.10	.70	-.04	.10	.09	.52
Wb Sense of well-being	.53	-.03	.36	.00	.08	.54	.71
Re Responsibility	.54	-.06	.20	-.02	-.42	-.05	.51
So Socialization	.34	-.19	.04	.14	-.09	.28	.26
Sc Self-control	.74	.04	-.20	.08	-.02	-.09	.60
To Tolerance	.68	.02	.43	.16	-.10	.07	.69
Gi Good impression	.56	.27	-.03	.10	.36	-.03	.53
Cm Communality	-.02	-.12	.04	-.04	-.11	.66	.47
Ac Achievement via conformance	.55	-.13	.33	.11	.30	.41	.70
Ai Achievement via independence	.49	.01	.32	-.03	.07	.13	.37
Ie Intellectual efficiency	.41	-.11	.41	.08	-.03	.51	.62
Py Psychological mindedness	.35	-.04	.11	.12	.14	.13	.19
Fx Flexibility	-.08	.28	-.01	-.11	-.20	.09	.15
Fe Femininity	-.10	-.06	-.06	-.10	-.39	.02	.18
Sentence Completion Change							
AS Active-soft	.27	.22	.15	.04	.39	.16	.32
AF Active-firm	-.15	-.04	.19	.05	.43	-.01	.25
PS Passive-soft	.01	.00	-.34	-.22	-.52	.23	.49
PF Passive-firm	-.11	-.16	.07	.08	-.43	-.47	.45
Client Ratings of Change							
Symptoms and complaints	.16	.07	.17	.77	.05	-.09	.66
Understanding of self	.04	.44	.27	.41	.09	.01	.44
Feeling and outlook on life	.13	.24	.08	.79	.04	.14	.73
Overall rating	.06	.11	.02	.85	.17	.02	.77
Therapist Ratings of Change							
Symptoms and complaints	-.02	.73	-.10	.18	.19	-.04	.61
Understanding of self	.08	.74	.12	-.01	-.03	.05	.60
Feeling and outlook on life	-.05	.64	-.05	.30	.24	-.25	.62
Overall rating	.01	.81	-.07	.20	.02	-.17	.73

rating of change in feeling. Surprise was expressed when the therapist ratings seemed largely independent of the other measures of change, but such surprise is even more appropriate in regard to the client ratings because the test score changes (especially the CPI) represent a kind of client rating in themselves. As was mentioned in discussing Factor A, the test score changes represent differences in performance at two different times and thus are sensitive to changes that the client may not be aware of at the end of therapy, since his rating of change depends on memory of his earlier state. It is interesting to note that the client's rating of change in understanding has only a moderate loading on this factor, suggesting that improved understanding is not so important in determining the client's evaluation of the outcome of therapy. Going back to Factor B, the therapist rating factor, for a moment, it is interesting that the therapist's rating of change in feeling has the lowest loading among the therapist ratings and that it does have a small loading on Factor D. This suggests that part of the lack of relationship between the therapist and client ratings may be due to a difference in values rather than complete lack of agreement on what changes have occurred. The therapists seem to consider self-understanding more important in their concept of improvement than does the client and conversely the client seems to value change in the way he feels somewhat more highly than does the therapist. Change in presenting symptoms seems to be about equally important for therapist and client.

Factor E is defined mainly by loadings of the sentence completion scores and seems to represent change in the ratio of active to passive solutions to the conflict presented in the sentence stem. The negative loading for Fe and the positive loading for Gi is consistent with the tentative interpretation of those sentence completion scores, but the negative loading of Re does not seem to fit into any reasonable interpretation.

Factor F has Cm, Ie, Wb, Ac, and PF as high loading tests. Since Cm, the highest loading variable, is a rare response scale and Wb, the second highest loading variable, is a response distortion scale, perhaps this factor represents change in the care taken in re-

sponding to the test items. This would be consistent with the negative loading for PF since the PF sentence completions represent easy avoidance of a conflict situation and are fairly obviously not desirable responses. This interpretation is supported by the fact that on the profile presented by Gough (1957) of CPI items answered at random plotted on the male norms, the four scales deviating most from the mean are the Cm, Ie, Wb, and Ac scales, and the rank order of their deviation is the same as the rank order of their factor loadings on the present factor.

Nature of the test score change factors. One question that is of interest is whether or not the factors identified in the test score changes are the same as those that would be isolated from the original test scores. An unpublished factor analysis of the 18 CPI scales done independently on different Ss was compared with the two CPI factors of the present study. Factor A is practically identical to the first factor of the CPI analysis and the correlation between factor loadings based on the 18 CPI scales is .90. Factor C is very similar to the second factor of the CPI analysis with a correlation between factor loadings of .91. Thus, there is clear-cut evidence that the two CPI change factors represent changes in the same influences that are responsible for covariance among the CPI scales in the first place.

Correlations between the two CPI change factors (A and C) and the corresponding factors in the pretests were estimated by averaging the correlations between the pretest scores and the corresponding change scores for the scales with the highest loadings for the factor using the z transformation. These correlations were $-.39$ for Factor A and $-.44$ for Factor C. This indicates that the lower (more maladjusted) the original score the greater the change tends to be. Since most of the Ss in this study had initial scores below the mean for college students, these correlations could be interpreted as indicating regression toward the mean due to unreliability of the test scores. Alternatively these negative correlations could be interpreted as indicating the greater tendency of the more maladjusted cases to improve, either because of therapy or spontaneously. The same regression effect

is present in the sentence completion scores as is indicated by the estimated correlation of $-.46$ between pretest score and change score on Factor E.

CHANGES OCCURRING WITH THERAPY

The above factor analysis has indicated the dimensions along which change occurs. However, it does not tell us whether or not the mean changes in the Ss receiving therapy are different from zero nor does it tell us whether or not these changes have been brought about or facilitated by the therapy.

Evaluation of Mean Change

The mean change in the salient variables for each of the factors is shown in Table 3. Every high loading variable for Factors A, B, C, and D shows significant change in a direction that would be considered to reflect improvement. The patterns of change for Factors E and F are not so clear-cut. Three of the four variables loading on Factor E show significant change but one of these, Re, is in the wrong direction. Only half of the high loading variables on Factor F show significant change, and this may be due to the loadings of these scales on other factors. Thus, Factors E and F cannot be said to show consistent changes.

These results show that the 75 Ss, all of whom received therapy, when considered as a group show definite improvement on the two CPI factors and on the therapist and the client rating factors.

Changes Related to the Therapy

Although the above results show significant improvement on four of six change factors, they do not indicate whether or not the improvement was produced by the therapy. There are, however, some comparisons within the present data that bear on this important question. One possible indication of the effect of therapy on the various change factors would be the correlation of the factors with the number of interviews on the assumption that the larger the number of interviews, the greater should be the therapeutic effect. These correlations can only be suggestive, however, since the number of interviews is

TABLE 3
SIGNIFICANCE TESTS OF CHANGE SCORES OF THE
SALIENT VARIABLES ON EACH FACTOR

Variable	Mean ^a	SD	<i>t</i> ^b
Factor A			
Wb	2.12	5.05	3.64**
Re	1.16	3.02	3.32**
Sc	1.24	4.70	2.62*
To	1.68	4.14	3.51**
Gi	1.21	4.68	2.18*
Ac	1.64	4.01	3.54**
Factor B (therapist ratings)			
Symptoms	3.27	.79	13.90**
Understanding	2.96	.56	12.20**
Feeling	3.21	.89	11.73**
Overall	3.17	.74	13.70**
Factor C			
Do	2.33	4.70	4.29**
Cs	1.93	3.60	4.64**
Sy	2.05	3.96	4.48**
Sp	2.89	5.26	4.75**
Sa	1.05	2.95	3.08**
Factor D (client ratings)			
Symptoms	3.43	.84	14.70**
Feeling	3.35	.98	11.92**
Overall	3.37	.85	13.95**
Factor E			
PS ^c	-2.47	9.42	2.27*
PF ^c	-0.35	9.70	0.31
Re ^c	1.16	3.02	-3.32**
AF	3.32	9.45	3.02**
Factor F			
Wb	2.12	5.05	3.63**
Cm	0.13	3.01	0.37
Ie	1.63	4.93	2.86**
PF ^c	-0.35	9.70	0.32

^a For test variables scores were obtained by subtracting pretest raw score from posttest raw score.

^b Test of significance of difference of mean from zero. In the case of ratings (Factors B and D) *t* test is of difference from a rating of no change.

^c These variables had negative loadings on the respective factors.

* Significant at .05 level.

** Significant at .01 level.

confounded with time between tests for the test score changes and probably is confounded with the placebo effect for the ratings. Also there is probably a nonlinear relationship between number of interviews and success as has been suggested by Cartwright (1955). Nevertheless, there was considerable variance

in number of interviews (mean = 14.7, $SD = 9.2$) and correlations between each factor and number of interviews were estimated by averaging the correlations of the highest loading variables with number of interviews using the z transformation. These combined correlations are shown in Table 4. With correlations based on 75 cases the correlation with Factor B (therapist rating) is the only one significant.

Another indication of the amount of therapy received might be based on the skill of the therapist. There were 41 therapists, each of whom was carefully supervised. Correlations between supervisors' rating of therapeutic ability of the therapist and the various change factors were estimated by the procedure used above. These correlations are shown in Table 4. Rated skill of the therapist is significantly related to both the therapist and client rating of change. Some contamination of the supervisors' rating is a possibility since the supervisor's judgment of the therapist may have been influenced by the progress of the case, but this artifact is attenuated by the fact that each of the therapists carried several cases. Even though all the cases were not included in this study, they were all considered by the supervisor in making his rating. Number of interviews was correlated a borderline significant .21 with supervisors' rating of skill of the therapist.

COMPARISON WITH A CONTROL GROUP

On the basis of the above approximations it appears that the therapist and client ratings are the only measures of change that are

TABLE 4

CORRELATIONS OF THE CHANGE FACTOR SCORES WITH
NUMBER OF INTERVIEWS AND SKILL OF THERAPIST

Change Score	Correlation with Number of Interviews	Correlation with Skill of Therapist
Factor A	.01	.09
Factor B	.29**	.26**
Factor C	.02	.07
Factor D	.15	.27**
Factor E	.05	.16
Factor F	.09	.01

** Significant at .01 level with one-tailed test.

influenced by the psychotherapy. However, comparisons within a therapy group leave much to be desired as a means of evaluating therapy, and data for a nontherapy control group was collected to serve as a baseline for the evaluation of changes in Factors A, C, and D (three of the four factors showing significant changes in the therapy group).

Considerations in Selecting Control Subjects

One of the most difficult problems in the evaluation of the results of psychotherapy has been the securing of adequate untreated control cases. Many investigators have omitted controls entirely. Some, such as Barron and Leary (1955), have used waiting list cases, and others, such as Rogers and Dymond (1954), have used a normal group of non-clinic Ss.

Although these two studies represent the most adequate controls available for changes in self-ratings with therapy, some objections can be raised to both. Barron and Leary have pointed out that there may be considerable therapeutic effect in an initial intake interview and in being on a clinic list. They also suggest that self-rating tests may be affected by the relationship to the clinic and therapist, and these relationships are quite different for clinic and waiting list controls at the time of the posttest. The nonclinic control group used by Rogers and Dymond has the disadvantage of not being comparable to the therapy group in terms of maladjustment. On both of their self-rating measures, the self-ideal correlation and the Q sort adjustment score, the nonclinic control group achieved higher mean scores on the pretest than any mean score ever achieved by the therapy group. Thus, if the tests have any ceiling effect, the control group will be more affected than the therapy group.

In the present study an attempt was made to secure a nonclinic control group that was comparable to the therapy group in a number of important respects. The control group used in this study had the following characteristics: (a) each therapy case was matched with a corresponding control case for sex of S and for initial scores on the criterion test (the CPI), (b) time between pretest and posttest for each control case was equivalent

to the time taken by therapy in the corresponding therapy case, (c) both therapy and control groups were drawn from the same student population which is fairly homogeneous with regard to age, (d) the control Ss were told at the time of the posttest that the effects of a course in which they were enrolled, introductory psychology; were being studied. This was an attempt to produce a similar test-taking attitude to that of the therapy Ss who knew that therapy was being evaluated when they took the posttest.

Procedure for Obtaining Control Data

Data for the therapy cases were collected over a two-year period. During the second year of the study a control group was obtained for the 42 therapy cases obtained during the first year. No control cases for the second-year therapy Ss were collected, so the *N* for all control group comparisons is 42.

Early in the fall semester of the second year of the study, 183 introductory psychology students signed up to participate in an experiment, which was a course requirement, and were given the CPI. The tests were administered in small groups. The CPI profiles for these 183 Ss were plotted on transparent paper and matched visually with the profiles of the therapy Ss. The control *S* of the same sex and with the most similar CPI profile was selected as the matched control case for each of the 42 therapy Ss. In those instances where several control Ss had similar profiles to a given therapy *S*, the one with the smallest *D* score (Cronbach & Gleser, 1953) was selected. This procedure yielded reasonably good matching in every case. Since the therapy and control groups were to be compared on the two CPI change factors it would have been desirable to match the groups on the basis of factor scores rather than profiles. However, the control group had to be collected before the factor analysis was done so this was not possible. Reasonably good matching on fac-

TABLE 5
COMPARISON OF INITIAL CPI FACTOR SCORES
OF THERAPY AND CONTROL GROUPS

Factor	Therapy		Control		<i>t</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>	
A	47.2	9.7	48.9	8.8	0.87
C	49.8	9.1	52.3	9.5	1.30

tor scores was obtained by the procedure used. The mean pretest factor scores for Factors A and C (the two CPI factors) of the therapy and control groups are shown in Table 5. Although fairly good matching of therapy and control pairs was obtained, this table shows a nonsignificant tendency for the control group to have higher scores.

The 42 matched control Ss were called in individually for retesting after a period of time equal to that between pretest and posttest for the corresponding therapy *S*. At this time the control Ss were told that the *Es* were evaluating the effectiveness of the introductory psychology course in bringing about personality changes. They were retested with the CPI and were given a slightly altered form of the rating scale shown in Table 1. References to the clinic were changed to make them appropriate to the introductory psychology course. Throughout the testing of the control group an attempt was made to structure the testing situation in a manner as similar to that of the therapy group as possible.

Factor scores for the two CPI factors were obtained by calculating the mean standard score change obtained from the CPI profile form for the highest loading variables on each factor. Thus, the factor score for Factor A is the mean standard score change for Scales Wb, Re, Sc, To, Gi, and Ac; and for Factor C is the mean standard score change for Scales Do, Cs, Sy, Sp, and Sa. The factor score for Factor D, client rating, is the mean rating for Scales 1, 3, and 4 of the rating scale shown in Table 1.

TABLE 6
COMPARISON OF THERAPY AND CONTROL GROUPS ON CHANGE FACTORS A, C, AND D

Factor	Therapy			Control			Difference between therapy and control (<i>t</i>)	
	Mean	<i>SD</i>	<i>t</i> ^a	Mean	<i>SD</i>	<i>t</i> ^a	Variances ^b	Means ^b
A	3.26	6.35	3.33**	4.07	4.79	5.52**	1.83	-0.56
C	5.14	7.58	4.39**	1.84	5.66	2.10*	1.89	2.38**
D	3.50	0.67	19.35**	2.73	0.53	11.90**	1.64	6.18**

^a Test of significance from zero or from rating of no change in the case of Factor D.

^b Test of significance of differences between variances and between means of the change scores of therapy and control groups taking into account the correlation of change scores brought about by the matching of therapy and control cases.

* Significant at .05 level.

** Significant at .01 level.

TABLE 7
COMPARISON OF THERAPY AND CONTROL CHANGE SCORES ON
INDIVIDUAL CLIENT RATINGS AND CPI SCALES

Scale	Therapy (N = 42)			Control (N = 42)			Test of difference (<i>t</i>) between therapy and control	
	Mean	SD	<i>t</i> ^a	Mean	SD	<i>t</i> ^a	Variances ^b	Means ^b
Factor A								
Wb Sense of well-being	2.64	5.74	2.98**	2.69	3.31	5.27**	3.68**	-0.48
Re Responsibility	0.60	3.07	1.26	1.19	4.42	1.74	-2.38**	-1.20
Sc Self-control	0.76	4.53	1.09	3.81	6.01	4.11**	-1.83	-2.08*
To Tolerance	1.88	4.66	2.62*	-3.17	6.45	-3.18**	-2.11*	3.38**
Gi Good impression	1.19	4.43	1.74	6.81	5.95	7.42**	-1.90	-4.66**
Ac Achievement via conformance	1.57	4.32	2.36*	2.26	3.71	3.95**	0.99	-0.78
Factor C								
Do Dominance	2.88	5.20	3.59**	1.59	5.32	1.94	-0.14	1.20
Cs Capacity for status	2.07	4.08	3.29**	0.24	3.25	0.48	1.56	3.08**
Sy Sociability	2.40	4.48	3.47**	1.09	4.18	1.69	0.44	1.31
Sp Social presence	3.71	5.52	4.36**	1.62	4.80	2.19*	0.90	1.80
Sa Self-acceptance	1.26	3.16	2.58*	0.14	3.32	0.27	-0.30	1.60
Factor D (client ratings)								
Symptoms	3.5	.55	17.62**	2.7	.47	9.64**	1.59	5.66**
Feeling	3.5	.79	12.30**	2.6	.53	7.28**	1.93	4.70**
Overall	3.5	.55	17.62**	2.9	.43	13.53**	1.89	3.74**
Not Scored on Any Factor								
CPI Scales								
So Socialization	0.60	4.27	0.90	1.88	5.34	2.28*	-1.45	-1.27
Cm Communalty	0.48	2.91	1.06	0.60	2.00	1.93	2.52	-1.10
Ai Achievement via independence	1.62	3.53	2.97**	1.24	3.54	2.27*	-0.01	0.72
Ie Intellectual efficiency	1.67	5.06	2.14*	1.60	4.86	2.13*	-.28	0.33
Py Psychological mindedness	0.26	3.44	0.49	0.98	2.31	2.75**	2.64*	-1.15
Fx Flexibility	0.60	2.35	1.64	-0.02	2.85	-0.05	-1.26	1.22
Fe Femininity	-0.50	2.86	-1.13	-0.19	2.16	-0.57	1.80	-0.51
Client Rating								
Understanding of Self	3.2	.65	11.95**	3.0	.34	19.02**	2.78**	1.56

^a Test of significance of difference from zero, or from a rating of no change in the case of client ratings.

^b Test of significance of difference between variances and between means of the change scores of therapy and control groups taking into account the correlation of change scores brought about by the matching of therapy and control cases.

* Significant at .05 level.

** Significant at .01 level.

Comparison of Therapy and Control Groups

The change scores for therapy and control groups on the three factors to be compared are shown in Table 6. This table shows that both therapy and control groups show significant changes on all three factors, all in the direction of improvement. When therapy and control groups are compared the therapy group shows significantly greater improvement than the control group on Factors C and D.

Other studies (such as Cartwright, 1956) have found therapy groups to have greater variance on change measures than control groups. This is also true of the present therapy and control groups on all three factors, but these differences are not statistically significant unless one is willing to use a one-tailed test for this comparison.

A comparison of therapy and control groups on the individual scales is shown in Table 7. Here it can be seen that the significant dif-

ferences between therapy and control groups on Factors C and D are due to consistent differences in all of the scales loading on these factors. This supports the contention that it is on the underlying dimension common to all scales loading on the factor that change occurs. This consistency is not found in the change scores loading on Factor A. The control group shows greater change than the experimental group on most of the scales loading on Factor A, but shows a significant change in the wrong direction on one scale, To. The inconsistent performance of the To change scores in the control group is puzzling, but a recheck of the data indicates that it is not due to recording or calculating error. The change scores with no high loadings on any factor (for the most part scales with low communalities in the factor analysis) show no significant differences between therapy and control groups. This suggests that the dimensions in the measures used in this study that are affected by psychotherapy are pretty well accounted for by the factors isolated.

Some Implications for Research Design

It may be appropriate to mention at this point the great clarification of the therapy and control comparison that was brought about by the factor analysis. The CPI data shown in Table 7 were available before the factor analysis was done, and without the organizing influence of the factors, the data showed the confusing result of the therapy group improving more on Cs and To and the control group improving more on Sc and Gi. Since the number of comparisons made was relatively large, it was difficult to evaluate the statistical reliability of this pattern of differences. However, when the scales are grouped according to their factor loadings as in Table 7, an orderly pattern emerges.⁴

The control group changes shown in Table 7 indicate the necessity of using a control group in the evaluation of psychotherapy change. It cannot be said on the basis of the present

data whether the control group changes are due to a general optimism leading to improved self-ratings or whether they represent regression effects due to the selection of the more maladjusted cases by the matching process. However, the marked tendency of the therapy cases with the lower test scores to show the greatest improvement, which was discussed above, suggests that it is necessary for an adequate control group to be matched with the therapy group on the criterion measures to be used.

DISCUSSION

The above results appear to show that psychotherapy is effective in producing personality changes that are reflected in changes on certain CPI scales and in behavior discernible by both therapist and client. However, the factor analysis has shown that these changes occur along three independent dimensions. Most investigators seem to assume that psychotherapy has a unitary effect in improving the adjustment of the client, and that lack of agreement among criteria is due to error in measurement of some sort. The present findings can be interpreted in a way that is consistent with this assumption of unidimensional change. Therapist and client ratings and CPI measures of sociability may all be sensitive to a unitary change in adjustment brought about by therapy, but each of these types of measures may in addition be affected by special biases or other errors which are responsible for the finding of independent factors. If this were the case, one would expect to find some correlation between the change scores on the three factors due to the common influence of the personality change in a group receiving therapy. An inspection of the correlation matrix of change scores reveals that client ratings had a low positive relationship with both the test score changes loading on Factor C and the therapist ratings. But the therapist ratings and test score changes were not correlated. However, the client ratings show similar correlations with the test score changes loading on Factor A and with the other scores that do not change with therapy. This suggests that the relationship between client ratings and the other variables is due to common biasing effects rather than the influence

⁴ Harrison Gough (personal communication) has pointed out that the grouping of CPI scales brought about by the factor analysis was already implicit in the four clusters of scales described in the manual. The first cluster is very similar to our Factor D and the second is very similar to our Factor A.

of a general psychotherapy change. The alternative interpretation of the present results is that the effects of psychotherapy are multidimensional. In view of the lack of correlation among change scores shown to be due to therapy, this interpretation seems somewhat more consistent with the data.

If the effects of psychotherapy are multidimensional, the psychotherapy itself must have several independent aspects which have differential effects on the various criterion measures. For example, if therapist ratings and test score changes both reflect improvement due to therapy, and yet they are uncorrelated in a group receiving therapy, they must be due to different aspects of the therapy.

The CPI change (Factor C) differentiates therapy and control groups, but does not appear to be related to either amount or intensity of the therapy as represented by number of interviews and skill of therapist. This suggests that the mere fact of having entered into a therapeutic relationship is sufficient to produce this change. The first association one is likely to have to this is that perhaps the CPI change is a manifestation of the tendency to present oneself as in need of help at the beginning and as not needing help at the end of therapy (the old hello-goodbye effect). There are, however, two findings that suggest that the CPI changes represented by Factor C are not due to increased tendency to present oneself in the best light at the end of therapy. The first is that therapy and control groups did not differ on Factor A, which is the factor that should be most sensitive to changes in tendency to present oneself in a good light. The second is the *Es*' nonquantitative observation that the therapy *Ss* did not feel any particular obligation to the clinic. The therapy group was much more difficult to schedule for posttests than the control group and many more of the therapy *Ss* missed posttest appointments and expressed resentment at the imposition than was the case with the generally cooperative control *Ss*.

Thus, the CPI change with therapy seems on the one hand to be an attitude change brought about by simply being in therapy, while on the other hand, it does not appear to be simply a change in tendency to present

oneself in the best light. Perhaps the CPI differences between therapy and control groups are due to differences in the changes expected by the *Ss* to accrue from therapy and introductory psychology. Or, alternately, perhaps they reflect genuine personality changes brought about by therapy. The latter interpretation would necessitate the assumption that the supervisors' ratings do not correspond to the therapists' ability to produce this particular change.

Therapist and client ratings were both related to the skill of the therapist as rated by the supervisor, and the difference between the correlation of these two variables with number of interviews was not significant. They were also significantly correlated with each other. (Correlations ranged from .23 to .36 on corresponding scales.) Thus, the assumption that they are reflecting independent changes is not as well justified as in the case of the CPI changes. The lack of higher correlation between them is probably due to different biases inherent in the different points of view. For example, data not presented in this study indicate that the therapists tended to rate as improved those with high (well adjusted) initial CPI scores, suggesting that their rating of change may be influenced by the general level of adjustment, while this was not observed in the client ratings.

One possible interpretation of the relationship of client and therapist ratings with skill of therapist is that therapist, client, and supervisor may all be reacting to the ability of the therapist to form pleasant, lasting relationships. However, if this were the case the clients would be expected to show evidence of presenting themselves in a good light in making the favorable ratings that produced the very significant difference from the control group on Factor D. Evidence discussed above suggests that this was not the case, thus it appears more likely that both therapist and client are rating real personality changes in the client.

SUMMARY

A number of measures of change with psychotherapy including several therapist and client ratings and change scores for 18 CPI scales were factor analyzed, using 75 therapy

cases as subjects. The six factors obtained from this analysis were largely identified by the measuring instruments used. Of the six factors four showed mean changes representing significant improvement. The effects of therapy on the various change factor scores were studied by comparison with a matched control group and by correlating the change factor scores with number of interviews and with rated skill of therapist. These comparisons indicated that psychotherapy has significant effects as measured by therapist and client ratings of change and by change in a group of CPI scales reflecting poise and interpersonal effectiveness.

REFERENCES

- BARRON, F., & LEARY, T. F. Changes in psychoneurotic patients with and without psychotherapy. *J. consult. Psychol.*, 1955, 19, 239-245.
- CARTWRIGHT, D. S. Success in psychotherapy as a function of certain actuarial variables. *J. consult. Psychol.*, 1955, 19, 357-363.
- CARTWRIGHT, D. S. Note on "Changes in psychoneurotic patients with and without psychotherapy." *J. consult. Psychol.*, 1956, 20, 403-404.
- CARTWRIGHT, D. S., & ROTH, I. Success and satisfaction in psychotherapy. *J. clin. Psychol.*, 1957, 13, 20-26.
- CRONBACH, L. J., & GLESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, 50, 456-473.
- EBNER, E. An investigation of a method of scoring reconciliations of contradictions. Unpublished master's thesis, Purdue Univer., 1958.
- GIBSON, R. L., SNYDER, W. V., & RAY, W. S. Factor analysis of measures of change following client-centered therapy. *J. counsel. Psychol.*, 1955, 2, 83-90.
- GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto: Consulting Psychol. Press, 1957.
- ROGERS, C. R., & DYMOND, ROSALIND, F. *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.
- ROSENBERG, A. L. The development of an instrument for assessing the modes of reconciling contradictions. Unpublished master's thesis, Purdue Univer., 1957.

(Received August 3, 1959)

A CROSS-VALIDATION OF THE HOUSE-TREE-PERSON DRAWING INDICES PREDICTING HOSPITAL DISCHARGE OF TUBERCULOSIS PATIENTS

ALDO SANTORUM

Veterans Administration Center, Martinsburg, West Virginia

In an earlier article, Vernier, Whiting, and Meltzer (1955) reported certain signs obtained from the house and person drawings of the House-Tree-Person (H-T-P) Test which differentiate significantly the MHB (maximum hospital benefit) discharged patient from the AMA (against medical advice) discharged patient upon his admission to a Tuberculosis Service. They found significant differences between MHB and AMA groups on 7 of 14 test signs selected for analysis (see Table 1, Original Data). It was concluded by the authors that:

Despite the small samples available for analysis, a sufficient number of reliable differences in test scores were found to permit accurate prediction for individual patients. The data for one of the tests, the drawing of the house, justified the development of a specific index.

PROBLEM

Because of the theoretical implications ascribed by the authors to the predictive signs, and because of their possible use in planning therapeutic programs for TB patients, the need for cross-validating the original results was deemed essential. The present study, then, represents an attempt to provide cross-validation information on comparable groups of TB patients. In fact, our second cross-validation sample represents, in part, the same patient population from which their original sample was drawn.

METHOD AND RESULTS

H-T-P Test protocols were obtained from a total of 397 patients admitted to our Tuberculosis Service during the period of December

1954 through June 1956. For purposes of analysis 397 "house" drawings and 383 "person" drawings were available. One hundred thirty-six of these patients received eventual AMA discharges, while 261 patients were discharged MHB. Because the N s of both groups were greater than 100, the formula recommended by McNemar (1949, p. 76) was utilized to test the significance of the differences between groups on each of the 14 test signs.

As noted in Table 1 (First Cross-Validation Sample) none of the differences between the groups reported significant in the earlier study was found in this current analysis. Indeed, a reversal in one of the signs, significant at the .05 level, was noted in the current sample. The findings suggested the need for further corroborative efforts.

Consequently, the H-T-P protocols of a second sample of TB patients admitted during the period January 1953 to January 1954 were studied. All available patients receiving either AMA ($N = 74$) or MHB ($N = 77$) discharges were used. With N s below 100 in both groups, McNemar's (1949, pp. 76-77) recommended formula for "small sample" differences between proportions, and the rule-of-thumb criterion which indicates when it is safe to compare groups by the D/σ_D technique were used. Again, as noted in Table 1 (Second Cross-Validation Sample) no significant differences between the AMA and MHB test protocols were found.

As a further step, the statistical findings of the original study were re-evaluated. The formula for small sample differences and the rule-of-thumb criterion again were employed. Significant differences between the two groups

TABLE 1
COMPARISON OF AMA AND MHB GROUPS ON THE HOUSE-TREE-PERSON TEST

Test Signs	Original Data			Reanalysis of Original Data	First Cross-Validation Sample		Second Cross-Validation Sample	
	AMA (N = 50)	MHB (N = 30)			AMA (N = 136)	MHB (N = 261)	AMA (N = 74)	MHB (N = 77)
	%	%	CR	CR	%	%	%	%
Door on left side of page only	6	27	2.3*	2.64 ^b	25	22	14	26
Door details, 2 or more present	24	30	0.6	0.6	15	18	24	23
Window detail drawn as +	62	23	3.9**	3.39**	40	31	34	22
Number of windows 4 or more	32	53	1.8	1.85	38	35	36	47
Smoke from chimney	26	0	4.3**	3.05 ^b	21	13	22	18
Walk present	20	10	1.2	1.17	13	8	12	22
Steps drawn	18	27	1.0	0.96	19	16	28	22
Road indicated	12	0	2.4*	1.97 ^b	2	1	15	22
Shrubby and/or trees drawn	2	10	2.0	1.5	3	4	4	9 ^b
Small side of house on right side of page	44	20	2.4*	2.17*	4	9 ^a	7	5 ^b
Ground line drawn	22	3	2.7**	2.30 ^b	10 (N = 134)	11 (N = 249)	18	12
Head facing to right of page	20	7	1.6	1.57	13	16	15	18
Neck either omitted or drawn wider than long	50	37	1.2	1.13	49	51	39	43
Hands omitted	44	23	2.1*	1.91	54	57	26	29

* Significant at .05 level.

** Significant at .01 level.

^a Reversed.

^b These cannot be safely interpreted as significant.

were found on six of the seven test signs originally reported as significant. However, as noted in the Reanalysis of Original Data section of Table 1, four of these differences did not meet the rule-of-thumb criterion, suggesting their very limited predictive value. The two remaining signs, which were reported as significantly different, did fulfill the rule-of-thumb criterion. However, as noted earlier, these significant differences were not confirmed in our two cross-validation samples.

DISCUSSION

It is difficult to account for the large discrepancies between results of the original study and the results obtained in the cross-validations. Investigation of the characteristics of the original and current patient samples revealed minimal differences between the groups with regard to sex, age, educational level, and personality traits. In addition, at least 50% of the second cross-validation sample included patients used in the original

study. Since the scoring method is rather clear-cut, the scoring of the data does not appear to be an area which contributed significantly to the obtained large discrepancies.

Had the authors of the original study applied McNemar's rule-of-thumb criterion a more cautious interpretation of the results of the original study would have been offered.

SUMMARY

The H-T-P protocols of two samples of patients who had obtained either MHB discharges or AMA discharges were analyzed in an effort to cross-validate the findings of an earlier study (Vernier, et al., 1955). None of the seven signs of the original study which were reported to differ significantly between the groups was found to hold up in the cross-validation.

The seven signs reported in the original study as significantly different were reanalyzed by appropriate statistical techniques; five of these signs were found to be statistically untenable.

The results of the present study suggest that the implications reported in the original study for projective test theory and for the problem of AMA and MHB behavior prediction should be utilized with extreme caution.

REFERENCES

- McNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
- VERNIER, CLAIRE M., WHITING, J. F., & MELTZER, M. L. Differential prediction of a specific behavior from three projective techniques. *J. consult. Psychol.*, 1955, 19, 175-182.

(Received September 10, 1959)

CHANGES IN RORSCHACH PERFORMANCE AND CLINICAL IMPROVEMENT IN SCHIZOPHRENIA¹

ROSALINE GOLDMAN²

Veterans Administration Regional Office, Baltimore, Maryland

The concept of withdrawal as descriptive of the behavior of the schizophrenic patient (Bleuler, 1916; Coon, 1944; Henderson & Gillespie, 1950; Masserman, 1946; Strecker, Ebaugh, & Ewalt, 1951) has been overshadowed by the controversies and differences in opinion about the consequences, etiology, and mechanisms of schizophrenia. The complex behavior known as withdrawal involves the total organism and manifests itself in a variety of ways, both in psychologic and physiologic behavior (Angyal, Freeman, & Hoskins, 1940). With clinical improvement following a schizophrenic episode, patients show greater interaction with others, more responsiveness to the environment, and greater expression of emotion (Greenblatt & Solomon, 1953). These changes are evidence of a return to "normal" behavior. We can use these changes in behavior to identify three areas in which withdrawal, so important in understanding schizophrenia, occurs. These can be grouped as the areas of (a) interpersonal relations, (b) attitudes towards or perception of the world, and (c) the emotional life. Each area may be related to the special interests of a different psychiatric school. The Sullivan school concerns itself particularly with withdrawal from people. Meyer (Cameron, 1944) and Campbell (1935) emphasize withdrawal from the world (environment). The Freudian school emphasizes withdrawal of emotions from external objects as libidinal energy is centered on or in the self.

In this study the assumption was made that the Rorschach test has satisfactory indices to reflect reactions in the areas of interpersonal relations, perception of the world and ability to deal with emotions, and that these indices are sensitive to changes within individuals. The Rorschach indices reflecting responses in the three areas were derived from Rorschach theory and were not empirically determined single scores. The indices were essentially those combinations of scores and ratios weighed and evaluated according to their theoretical significance by the clinician when he interprets a Rorschach protocol.

The Rorschach scores reflecting the empathy, rapport, and feeling toward others and which were called indices of interpersonal relations are (a) the Movement and Human responses ($M + H$) and (b) the ratio of form-color responses to the sum of color-form and pure color responses ($FC:CF + C$). The patient's perception of the world was reflected in four sets of Rorschach indices: (a) the perception of stimuli as others see them ($F + \%$), (b) conformity with the thinking of the group (the number of Popular responses), (c) evidence that the world is a source of fear and danger (responses with destructive and tension-laden content), and (d) distortion in the distance maintained by the individual between himself and the world (Rejections, Denials, Self-References). The expression and control of emotion were reflected in four sets of Rorschach indices: (a) the ratio of intellectually determined responses to the total number of responses ($F\%$), (b) the affective energy available for response to external stimuli (8-9-10%), (c) emotional maturity and control ($FC:CF:C$), and (d) the ratio of the sum of the emotional re-

¹ This study was carried out at the Boston Psychopathic Hospital while the writer was a USPHS Fellow. The complete report of the study is on file at Boston University Library, as the writer's unpublished dissertation of the same title (1955).

² The writer is indebted to Chester C. Bennett for his invaluable critical assistance.

sponses which are controlled to those responses where control is secondary or lacking ($FC + FY + FV:CF + C + YF + Y + VF + V$).

PROCEDURE

The general hypothesis tested was that patients who show clinical improvement have significantly greater incidences of changes in the direction of healthier Rorschach responses than patients who do not show clinical improvement. Changes in the direction of healthier responses as these are judged in Rorschach practice were taken as evidence of diminution of withdrawal.

The Rorschach protocol of each patient which was obtained when he was hospitalized as acutely ill was compared with his protocol obtained just preceding his departure from the hospital, regardless of condition at time of discharge. Predictions for the changes in each set of Rorschach indices accompanying changes in clinical condition were:

1. In the area of interpersonal relations, clinically improved patients would show changes in the direction of a greater number of healthy movement and human responses ($M + H$) and a relative increase in the ratio of FC responses to the sum of $CF + C$ responses.

2. In the area dealing with perception of the world, the clinically improved patient would show changes in the $F + %$ in the direction of the optimum range; the number of Popular responses would increase; Content would contain fewer signs of tension; and there would be fewer Rejections, Denials, Self-References.

3. In the area dealing with the expression and control of emotion, patients showing clinical improvement would show change in the direction of the optimum range in the $F\%$ and in the 8-9-10%; the $FC:CF:C$ distribution would be away from the "regressive" shift; and the form-dominated responses ($FC + FY + FV$) would increase relative to the responses where form is secondary or lacking ($CF + C + YF + Y + VF + V$).

Only cases diagnosed as having an acute schizophrenic reaction, regardless of type, and whose illnesses were of less than six months' duration, and who were between the ages of 16 and 40 were included in this study. This approximates the range Henderson and Gillespie (1950) give as the ages within which two-thirds of the schizophrenias have their onset. Patients with chronic alcoholism, epilepsy, mental defect, physical handicaps from birth, psychosurgery, and those who were already undergoing somatic therapy were excluded. The average age for the total group was 24.02 years; average education was 12.06 years of schooling. Only one patient in this study was over 35 years of age. The present hospitalization was his second, with a remission period of more than five years. Two other patients also had been hospitalized previously; and in each instance an interval of more than two years had elapsed between hospitalizations, during which time the patients had made economic and social ad-

justments at their premorbid levels. In all, the data were obtained from the Rorschach records of 45 cooperative acutely ill schizophrenic patients. The 45 cases were consecutive hospital admissions selected only on the basis of diagnosis, age, and recency of onset of illness. Each patient in this study was examined both when he was acutely ill and when he left the hospital either as improved or unimproved. Each individual was both an experimental subject (S) and his own control.

Since the general hypothesis of this study involved comparison of groups of improved and unimproved patients, changes in clinical condition had to be established. Two psychiatrists rated the patient, both on his admission to the hospital and again upon his departure from the hospital, for severity of illness. There was complete agreement between the psychiatrists that 12 of the patients showed no improvement in their clinical condition in the intervals between the administration of the two tests. These cases became the unimproved group. There was agreement in the psychiatric judgments that 14 patients showed "marked" improvement. The remaining 19 cases showed moderate or slight improvement in the opinions of both doctors. Thus, 33 cases were judged to show clinical improvement during their hospitalizations.

In order to compare these groups on the experimental data, hypotheses that the improved and unimproved groups came from the same population were tested. When the statistical tests were made for age, education, length of hospitalization, kind of therapy, and length of time between examinations, the null hypotheses could not be rejected and the assumption that all patients came from a single population was held tenable.

The changes in Rorschach indices between the first and second examination were the experimental data. A change indicating greater interest in others, or in the world, or better control of emotions was considered a positive change. If positive change was not shown, the case was classified as showing no change on a particular index. The direction of the change rather than the amount of change constituted the experimental data.

The positive changes for each index of withdrawal in each patient's pair of Rorschach protocols were analyzed separately and the incidences of positive changes for the improved and unimproved groups were compared. In addition, the changes in Rorschach indices for each area of withdrawal were combined and analyzed as a single score for incidence of multiple changes in the improved and unimproved groups. Three multiple incidence values were thus possible for the area dealing with the attitude toward others: changes in both the $M + H$ constellation and the $FC:CF:C$ ratio, changes in only one of these indices, or no change in either index. In each of the two other areas dealing with withdrawal (attitude toward the world, expression and control of emotion) there were five possible multiple incidence values. Using the chi square statistic (McNemar, 1949) the incidences of positive

TABLE 1
VALUES FOR SIGNIFICANCE OF CHANGES IN RORSCHACH INDICES OF WITHDRAWAL
IN CLINICALLY IMPROVED VS. UNIMPROVED CASES

Rorschach Index	All Improved Cases (<i>N</i> = 33)	Cases Showing Marked Improvement (<i>N</i> = 14)	Cases Showing Some Improvement (<i>N</i> = 19)
Attitude toward People			
<i>M</i> + <i>H</i>	<.04	<.03	<.15 ^a
<i>FC:CF</i> + <i>C</i>	<.04	<.04	^b
Combined Indices	<.01	<.01	<.05
Attitude toward Environment			
<i>F</i> + %	<.03	<.02	^b
Content	<.02	<.01	<.01
Popular	^b	^b	^b
Rejection-Denial	<.15 ^a	<.15 ^a	^b
Self-Reference			
Combined Indices	<.01	<.01	<.01
Expression and Control of Emotion			
<i>F</i> %	<.04	<.02	<.15 ^a
8-9-10%	<.04	<.02	<.15 ^a
<i>FC:CF:C</i>	<.01	<.01	<.02
<i>FC+FY+FV:CF+C+</i>	<.10 ^a	<.03	^b
<i>YF+Y+VF+V</i>			
Combined Indices	<.02	<.01	<.15 ^a

^a Trend toward significance.

^b Not significant.

Rorschach changes in the total group of improved cases (33 patients) were compared with the incidences of these changes in the unimproved group (12 cases). In the same way, the incidences of positive changes in the 14 cases showing marked improvement, and the 19 cases showing "some" improvement were compared with the incidences of these changes in the 12 cases showing no improvement (Table 1, Columns 2 and 3). A *p* value less than .05 was the level set in this study for rejection of a null hypothesis. Since only direction of change (positive, in the direction of "healthier" responses) was studied, the *p* values for the one-tailed test (Jones, 1952) were used.

RESULTS

The significance of the incidences of positive changes in the groups showing improvement when compared with the unimproved group for each index and the combined indices is given in Table 1. The greater discriminatory power of the combined Rorschach indices corroborates Rorschach theory and practice which holds that individual scores must be combined and patterned out for meaningful interpretation. Incidences of positive changes in the entire group of 33 im-

proved cases were significantly greater for 7 of the 10 individual Rorschach indices when compared with the incidences of positive changes in the 12 unimproved cases. The cases showing marked improvement showed significant changes on 8 of the 10 indices. In contrast to this, the cases showing only "some" improvement showed significant changes on only 2 of the 10 indices. The greater clarity and higher instances of positive changes in the markedly improved group suggest that "pure" cases give findings more significant than less definitive cases even though the number of cases in a sample is considerably reduced by using only the pure cases.

DISCUSSION

The Rorschach changes which most significantly differentiated between the improved and unimproved cases were seen in the reversals of the regressive shift in the *FC:CF:C* "ladder." The increase in the *FC* responses suggests that with clinical improvement the ego functions more effectively and there is a greater control over emotions. Use of the

qualitative changes in each patient's record, such as changes from minus to plus signs in a given score, may have contributed to the high significance consistently shown for the association between the reversal of the regressive shift and clinical improvement.

Cases evaluated as markedly improved clinically showed increased control of both emotional responsiveness and emotional tensions (increases in $FC + FY + FV$). The combined improved cases showed only a trend toward this change. The inference can be made that the more disturbing emotions indicated by shading and vista responses are especially difficult to deal with, since only the cases showing marked improvement can handle them effectively.

The decrease in pathological content which was found to be significant in all improved groups suggests that the world becomes less hostile or barren to the clinically improved patient regardless of the degree of improvement he shows. The significant positive changes in the $F + \%$, indicating greater ability to deal with reality, may be dynamically interlocked with the significant change found in decreased pathological Content. Together, these changes suggest that the world is "tested" more realistically by the patient because he is less afraid, and that he may be less afraid because of greater intellectual control and ego strength.

Two studies dealing with retest changes report negative findings when using Popular responses. Kelley, Margulies, and Barrera (1941) found fluctuations in the number of Popular responses given by amnesic patients when almost all other factors on retest were stable. Siegal (1948) found that Popular responses "occurred diffusely" in patients showing improvement as well as in those failing to show improvement. Molish (1951) found that three-fourths of the failures to give Popular responses by schizophrenic patients occurred on cards which usually elicit human percepts. In light of the relationship found in the present study between the $M + H$ index and clinical change, his findings suggest that we might have seen meaningful changes in Popular responses if the analysis had been made in terms of their content rather than in terms of their total number in a protocol.

Although prognostic signs were not the concern of this study, relevant data are available from the analysis of the initial test records of the improved and unimproved cases. These groups did not differ significantly, upon admission to the hospital, in the instances of healthy and unhealthy responses on any of the Rorschach indices used in this study. $F + \%$ was slightly more disturbed in the patients who showed marked improvement later, but this index showed no trend toward discriminating between those cases who later improved and those who did not improve when the degree of improvement was undifferentiated.

SUMMARY

To study changes in withdrawal which occur with clinical improvement, Rorschach indices of responsiveness to people, the environment, and to the patient's own affective life were selected. The indices reflecting activity in these three areas were derived from Rorschach theory and were not empirically determined scores. The indices were essentially those combinations of scores and ratios weighed and evaluated according to their theoretical significance by the clinician when he interprets a Rorschach protocol.

The Rorschach records of 45 cooperative acutely ill schizophrenic patients between the ages of 16 and 40 were studied. These cases were consecutive hospital admissions selected only on the basis of diagnosis, age, and recency of onset of illness. Each patient in this study was tested both when he was acutely ill and when he left the hospital as either improved or unimproved. Each individual was both an experimental subject and his own control; changes between first and second protocols were analyzed.

The findings in this study show that the Rorschach technique brings the functioning personality into focus so that the assumption that this instrument is sensitive to changes within an individual is justified. Combined indices were more significantly related to clinical improvement than the separate indices, except for the $FC:CF:C$ ratio. This is in accord with Rorschach theory and practice which hold that individual scores must be combined for meaningful interpretation of a protocol.

REFERENCES

- ANGYAL, A., FREEMAN, H., & HOSKINS, R. G. Physiologic aspects of schizophrenic withdrawal. *Arch. Neurol. Psychiat.*, 1940, 44, 621-626.
- BLEULER, E. *Textbook of psychiatry*. (Trans. by A. A. Brill) New York: Macmillan, 1924.
- CAMERON, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 2. New York: Ronald, 1944. Ch. 28.
- CAMPBELL, C. M. *Destiny and disease in mental disorders—with special reference to the schizophrenic psychoses*. New York: Norton, 1935.
- COON, G. P. Principal psychoses. In H. C. Solomon & P. I. Yakovlev (Eds.), *Manual of military neuropsychiatry*. Philadelphia: Saunders, 1944. Topic 23.
- GREENBLATT, M., & SOLOMON, H. C. *Frontal lobes and schizophrenia*. New York: Springer, 1953.
- HENDERSON, D., & GILLESPIE, R. D. *A text-book of psychiatry for students and practitioners*. (7th ed.) London: Oxford Univer. Press, 1950.
- JONES, L. V. Tests of hypotheses: One-sided vs. two-sided alternatives. *Psychol. Bull.*, 1952, 49, 43-46.
- KELLEY, D. M., MARGULIES, H., & BARRERA, S. E. The stability of the Rorschach method as demonstrated in electric convulsive therapy cases. *Rorschach res. Exch.*, 1941, 5, 35-43.
- MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
- MASSERMAN, J. *Principles of dynamic psychiatry*. Philadelphia: Saunders, 1946.
- MOLISH, H. B. The popular response in Rorschach records of normals, neurotics and schizophrenics. *Amer. J. Orthopsychiat.*, 1951, 21, 523-531.
- SIEGAL, M. G. The diagnostic and prognostic validity of the Rorschach test in a child guidance clinic. *Amer. J. Orthopsychiat.*, 1948, 18, 119-133.
- STRECKER, E. A., EBAUGH, F. C., & EWALT, J. R. *Practical clinical psychiatry*. (7th ed.) New York: Blakiston, 1951.

(Received September 14, 1959)

INTELLECTUAL FUNCTIONING IN TEMPORAL LOBE EPILEPSY¹

OSCAR A. PARSONS² AND DAVID E. KEMP

Duke University and Veterans Administration Hospital, Durham, North Carolina

Recent reports of impaired verbal abilities in psychomotor or temporal lobe epileptics have given new impetus to the continuing search for patterns of cognitive functioning associated with abnormal brain conditions. Quadfasel and Pruyser (1955) found that psychomotor epileptics had significantly lower Wechsler-Bellevue Verbal IQs in comparison to Performance IQs than did a group of *grand mal* epileptics. They further reported that memory for verbal material as measured by the Wechsler Memory Scale was significantly impaired for the psychomotor group while memory for designs was intact. Meyer and Jones (1957) confirmed the findings on the Wechsler-Bellevue in English patients and demonstrated that the deficit became greater subsequent to temporal lobe operation. Clinical evidence has mounted over many years to indicate that speech and communicative skills are associated with the temporal lobes, especially that of the left or dominant hemisphere. Since most psychomotor epileptics have left temporal lobe foci (Gibbs & Gibbs, 1952), the clinical and experimental findings of ver-

bal impairment in this group appear consistent.

The above results not only bear on the question of specificity and localization of brain functions but also seem to hold promise for the clinician who is faced with many perplexing diagnostic problems in connection with psychomotor epilepsy, e.g., differentiation from anxiety states, schizophrenia, psychopathic behavior, etc. However, before warrant may be given for generalization, the possible restrictions and conditions should be explored. One such restriction has already been raised by Meyer and Jones (1957), who found that only left temporal lobe epileptics manifested the verbal deficits; right temporal lobe subjects (Ss) did not. On the other hand, Quadfasel and Pruyser report the same Verbal-Performance discrepancy in both right and left temporal lobe epileptics. Examination of previous studies suggested that variables such as unilaterality vs. bilaterality of lesion, "pure" psychomotor vs. psychomotor mixed with generalized seizure patterns have been confounded. One purpose of the present study, then, is to examine those cognitive abilities represented by the Verbal and Performance Scale scores of the Wechsler Adult Intelligence Scale (WAIS) in a population characterized by strictly *unilateral* temporal lobe EEG foci and pure psychomotor seizure patterns.

A second aspect of the study is the exploration of possible differences in pattern of cognitive abilities as represented by subtest scoring on the Wechsler among psychomotor epileptics, generalized seizure (*grand mal*) epileptics, and control Ss. That such a pattern may exist for Ss with temporal lobe dysfunction is suggested by Milner's (1954) report of disturbances in visual comprehen-

¹ This report is based in part on a thesis submitted by the junior author in partial fulfillment of the degree of Bachelor of Arts with distinction at Duke University, under the direction of the senior author. The authors are grateful to Joseph B. Parker, Chief, Psychiatric Service, Durham VA Hospital; and Bernard Stotsky and Paul Daston, Chiefs, Psychological Service, Durham VA Hospital, for their aid in carrying out this study. This research was supported in part by a grant (B-1459) to the senior author from the National Institute of Neurological Diseases and Blindness, USPHS; and was carried out in conjunction with a larger VA study, *The Social and Emotional Adjustment of Epileptics*, under the direction of Joseph B. Parker, William Wilson, and Lever Stewart.

² Parsons is now at the Department of Psychiatry, Neurology, and Behavioral Sciences, University of Oklahoma Medical Center.

sion, as measured by Picture Arrangement type tests; in Reitan's (1957) proposition that the Similarities subtest of the Wechsler is particularly affected by temporal lobe damage and in Penfield's stimulation studies (Penfield & Jasper, 1954) which have led him to postulate that the temporal lobes have an important function in memory, comparison, and generalization. The specific pattern, then, may take the form of lower Similarities (generalization and comparison function), Information and Digit Span (remote and immediate memory), and Picture Arrangement (visual comprehension) for the psychomotor group in comparison to control (nonCNS damaged) Ss.

Again, however, relevant variables must be considered. Degree of personality disturbance, a variable often ignored in studies of the brain damaged, seems pertinent. It is recognized that psychomotor epileptics manifest more personality disorders than other epileptic groups (Gibbs & Gibbs, 1952) and that personality disturbances may affect cognitive functions (Wechsler, 1958). It would appear that a nonepileptic, nonCNS damaged control group, which was equated for degree of personality disturbance, would give the clearest test for patterns on the Wechsler associated with temporal lobe epilepsy. A generalized seizure group who had foci in the same area as the psychomotor group would provide a test for the uniqueness of any pattern found in the latter group. However, it might be more reasonable to expect that groups with similar localization of lesions (as defined by EEG foci) would manifest a similar pattern of cognitive difficulties. In summary, the first two predictions to be tested are:

1. Psychomotor and generalized seizure epileptics with EEG foci in the same general area (unilateral) have lower Verbal than Performance Scale scores on the WAIS.

2. These epileptic patients also have a pattern of cognitive functioning characterized by lower Information, Similarities, Digit Span, and Picture Arrangement subtest scores, in comparison to a group of nonepileptic controls equated for degree of personality disturbance.

A third purpose of this research is the investigation of the effect of degree of abnormal

EEG activity on the WAIS performance of the epileptic groups. Recent findings by Hovey and Kooi (1955) suggest that fluctuations in the level of responding (with consequent lowering of quality of response) on the Wechsler are related to the presence of subclinical paroxysmal EEG activity, the latter occurring in the absence of overt or clinical manifestations of seizures. It is anticipated that the relationships stated above will be more clearly established in those epileptics having a greater degree of abnormal EEG activity.

METHOD

Subjects

The Ss in this study were 46 male, right-handed patients of the Durham VA Hospital. Patients were placed into three groups, Psychomotor, Generalized, and Control, on the basis of criteria stated below. These three groups were equated for age, race, and education and were not significantly different on Full Scale WAIS IQs (Table 1).

Fifteen Ss were placed in the Psychomotor group on the basis of three criteria: (a) the presence in two or more electroencephalograms of interictal epileptiform discharge localized in *only one* of the temporal lobes (5 Ss with right temporal foci and 10 with left foci comprised this group); (b) the presence of seizures which were clearly epileptic and only psychomotor in nature, as determined by clinical observation and/or histories elicited from the patient and relatives or acquaintances familiar with the patient's seizures; and (c) an age of onset of epileptic symptoms of 20 years or older.

Sixteen Ss were selected for the Generalized group on the basis of type of seizure (generalized, tonic-clonic seizures) and the presence of focal epileptiform activity in two or more EEGs. Seven of these patients had psychomotor seizures in addition to their *grand mal* symptoms. The foci of this group was mixed, although predominately temporal lobe (14 out of 16 patients): 5 Ss had left temporal foci, 2 right temporal, 3 bitemporal, 2 left frontotemporal, 1 left parietal-temporal, 1 right temporal-occipital, 1 left frontocentral, and 1 right frontal pole. Onset of seizure symptoms also was after 20 years of age. The differences between the Psychomotor group and the Generalized group on mean age of onset and on duration of disorder are not significant (Table 1).

All epileptic patients were classified as to degree of abnormal EEG activity. "Abnormal" was defined as the presence of any of the following: spikes, sharp waves, sharp and slow waves complexes, multiple

³ The authors wish to thank William P. Wilson and Lever F. Stewart, certified electroencephalographers, who furnished the diagnostic information reported in this study. The EEG criteria were based on Penfield and Jasper (1954).

TABLE 1
DESCRIPTION OF THE CONTROL, PSYCHOMOTOR, AND GENERALIZED GROUPS

	Age	Education	Full Scale IQ	Age of Onset Symptoms	Duration of Disorder
Control (<i>N</i> = 15)					
\bar{X}	31.5	9.2	94.1	—	—
<i>SD</i>	6.3	3.6	13.4	—	—
Psychomotor (<i>N</i> = 15)					
\bar{X}	32.1	9.3	93.7	25.1	5.1
<i>SD</i>	10.4	5.9	13.6	10.0	4.2
Generalized (<i>N</i> = 16)					
\bar{X}	32.2	9.0	88.5	29.3	4.6
<i>SD</i>	6.5	3.4	11.3	7.8	3.0

Note.—None of the differences among groups is significant, as measured by *t* tests.

spike bursts, and spike and wave discharges. Records were rated on a three-point scale ranging from minimal activity (occasional or rare occurrence) through moderate to marked (abnormal activity occurring almost all the time) by a board electroencephalographer. In all other respects the neurological examination of these patients was negative.

The Control group consisted of 15 patients who had no evidence of CNS pathology and who were referred to the Psychological Service for psychodiagnostic testing. These Ss were equated with the Psychomotor group and Generalized groups on pertinent variables and for the degree and pattern of personality disturbance as manifested on the Minnesota Multiphasic Personality Inventory (MMPI). The MMPI *T* scores of the three groups, Control, Psychomotor, and Generalized, do not differ significantly (Table 2). The collective profile is marked by an elevated neurotic triad and psychasthenia and schizophrenia subscales.

Procedure

The Ss in the two epileptic groups were tested by the junior author on the WAIS, as part of a larger battery of tests. The WAIS was administered according to standard procedure (Wechsler, 1955).

Since another block design test was used in the larger battery, the Block Design subtest was omitted; therefore, the Performance Scale and Full Scale IQs reported for these Ss were prorated.⁴ The MMPI (Group Form) was administered to these Ss as part of the routine intake procedure of the Psychological Service.

Control Group Ss were administered the WAIS and the MMPI by various examiners from the psychological service.

RESULTS

The means and standard deviations for the Performance Scale IQs, Verbal Scale IQs, and Full Scale IQs are presented in Table 3. A Lindquist Type I analysis of variance of the Performance Scale and Verbal Scale IQs re-

⁴ As members of the general class of brain injured, epileptic Ss would be more likely to score *lower* on the Block Design than nonepileptic groups and thus reduce discrepancies between the expected lower Verbal than Performance scores. Thus, omission of the B-D should favor the predicted higher Performance Scale.

TABLE 2
MEAN MMPI *T* SCORES OF THE CONTROL, PSYCHOMOTOR, AND GENERALIZED GROUPS

	<i>L</i>	<i>F</i>	<i>K</i>	<i>Hs</i>	<i>D</i>	<i>Hy</i>	<i>Pd</i>	<i>Mf</i>	<i>Pa</i>	<i>Pt</i>	<i>Sc</i>	<i>Ma</i>	<i>Si</i>
Control	56	60	57	77	77	75	69	59	59	69	71	60	54
Psychomotor	53	61	57	72	75	73	71	51	65	71	75	63	53
Generalized	50	55	51	75	72	69	64	55	56	64	65	63	54

Note.—None of the differences among groups is significant as measured by *U* tests.

TABLE 3

MEAN PERFORMANCE IQ, VERBAL IQ, AND MEAN "PERFORMANCE MINUS VERBAL SCORE"
FOR CONTROL, PSYCHOMOTOR, AND GENERALIZED GROUPS

	Performance IQ	Verbal IQ	Performance-Verbal Score
Control			
\bar{X}	94.2	93.7	+4.5
SD	14.6	12.0	6.9
Psychomotor			
\bar{X}	91.5	96.0	-4.5
SD	13.8	14.3	10.6
Generalized			
\bar{X}	86.3	91.1	-4.8
SD	11.7	12.4	11.0

Analysis of Variance				
	df	MS	F	
Between groups	2	276.8	.93	ns
Error (B)	43	298.0		
Between trials	1	204.0	4.35	$p < .05$
T \times G	2	66.3	1.41	ns
Error (W)	43	46.9		

veals a significant trials effect. It is evident from inspection of the means that this difference lies in the two epileptic groups, both of whom have approximately the same discrepancy, i.e., a lower Performance than Verbal IQ which is opposite to the first prediction. Comparison of the left temporal and right temporal lesion cases in both epileptic groups indicates that the right temporal group ($N = 10$) have approximately the same Performance and Verbal IQs while the left temporal group ($N = 16$) manifests the discrepancy.

To test the second prediction, a difference score was obtained by subtracting the \bar{X} of

the scaled scores on Information, Similarities, Digit Span, and Picture Arrangement from the \bar{X} of the scaled scores for the remaining six subtests for each S. A positive score for the epileptic groups means a difference in the predicted direction. In the last column of Table 4 the difference score is presented for each group. The Generalized group differs significantly ($t = 2.17$, $p < .05$) from the Control group in the predicted direction, i.e., having a lower \bar{X} Information, Similarities, Digit Span, and Picture Arrangement score; the Psychomotor group does not. Examining the pattern of subtest scoring reveals that the weight of the difference between groups is

TABLE 4

MEAN WAIS SUBTEST SCORING OF THE CONTROL, PSYCHOMOTOR, AND GENERALIZED GROUPS

	Information	Comprehension	Arithmetic	Similarities	Digit Span	Vocabulary	Digit Symbol	Picture Completion	Picture Arrangement	Object Assembly	Difference* Score
Control	9.5	10.3	9.3	9.8	9.2	8.5	6.8	10.0	8.9	8.6	-.2
Psychomotor	9.1	10.5	10.1	9.1	8.7	8.2	6.5	9.7	7.6	9.0	+.2
Generalized	7.9	9.8	9.7	7.2**	8.6	8.3	6.0	9.4	7.0***	7.2	+.7

* Generalized group significantly different from Controls, $p < .05$.

** Generalized group significantly different from Controls, $p < .01$.

*** Generalized group different from Controls at $.10 < p < .05$.

TABLE 5
MEAN AGE, EDUCATION, PERFORMANCE, AND VERBAL IQS FOR CONTROL, MINIMAL,
AND MODERATE-SEVERE ACTIVITY GROUPS

	Age	Education	Performance IQ	Verbal IQ
Controls ($N = 15$)	31.5	9.2	94.2	93.7
Minimal ($N = 14$)	31.8	7.9	92.6	91.3
Moderate and Marked ($N = 17$)	36.1	9.9	85.6	95.6

carried by Similarities ($p < .01$) and to a lesser extent by Picture Arrangement ($.10 < p < .05$). Information and Digit Span are not significantly different.

The effect of degree of abnormal EEG activity was tested by comparing the fourteen Ss, from both the Psychomotor and Generalized groups, rated as having minimal activity, to the 17 Ss falling in the moderate and marked categories. In Table 5 the means for the control variables and Performance and Verbal IQs are noted for the three groups. It is evident that the Minimal group tends to be lower in education and the Moderate-Marked group is somewhat older. The latter variable, however, is not likely to produce an effect since both Verbal and Performance IQs are adjusted for age (Wechsler, 1958). In light of a near significant F ($.10 < p < .05$) for differences between the two abnormal activity groups on education and the finding of a correlation (r) of $+.59$ between education and Performance IQ, an analysis of covariance was performed (Edwards, 1950). The result of this analysis for the Performance IQ indicates a highly significant differ-

ence between groups, the Marked-Moderate group having a much lower Performance IQ than would be expected for their educational level in comparison to the Minimal group.

Analysis of covariance for the Verbal IQ, on the other hand, revealed no differences.

DISCUSSION

The finding that the epileptic groups have lower Performance than Verbal IQs contrary to the first prediction, raises a number of questions concerning generalization from previous findings. Lateralization of lesion could not explain these results since the discrepancy was more pronounced in the left temporal focus group as opposed to the right sided focus patients. Equating the Control group to the epileptic groups on degree and pattern of personality disturbance, while admittedly only an approximation, could have contributed to *between* group differences but not to the obtained *within* group differences in the epileptics. In seeking to illuminate the possible cause for the discrepancy between our results and others, three variables appear important.

First, the cortical functioning of our population of temporal lobe epileptics was probably less severely disturbed than other groups. In contrast to the present group, the Meyer and Jones (1957) epileptics were candidates for temporal lobe surgery. Quadfasel and Pruyser's (1955) patients had a longer duration of illness and earlier age of onset. While this might explain *no difference* between groups, it could hardly account for the significantly lower Performance Scale IQs.

Second, our population has a lower Full Scale IQ than the previous studies. To estimate the possible effect of this variable, the published data of Meyer and Jones (1957) were analyzed (preoperative scores). By taking the Full Scale IQ as the best overall meas-

TABLE 6
ANALYSIS OF COVARIANCE OF PERFORMANCE
AND VERBAL IQS IN ABNORMAL EEG
ACTIVITY GROUPS

Performance IQ	df	MS	F
T	29		
Within	28	54.11	
Adjusted Mean	1	1381.00	25.52**
Verbal IQ			
T	29		
Within	28	94.76	
Adjusted Mean	1	44.9	.474

** $p < .001$.

ure of the general level of intelligence, the left-sided lesion group of patients were divided at the median into High and Low IQ groups. The High IQ group ($N = 10$) averages 9.8 points lower on the Verbal Scale relative to Performance IQs while the Low IQ group ($N = 10$) had a .2 difference. This difference between groups is significant at the .05 level by U test. Since the patients used by Quadfasel and Pruyser (1955) had an even higher average IQ than Ss used by Meyer and Jones and the present study, it may be the verbal deficit reported in preoperated temporal lobe epileptics holds only in high IQ groups. The complexity of this issue is emphasized by the recent report of the preoperative Wechsler-Bellevue Scores for left and right sided temporal lobe lesion groups by Milner (1958). In this study, using comparatively high IQ groups, the preoperative Verbal and Performance IQs did not differ for the left-sided group, but the right-sided group had significantly lower Performance IQs!

The third variable is the difference in test instrument. The previous studies used the Wechsler-Bellevue, while the present study employed the WAIS. Studies cited by Wechsler (1958, pp. 114-117) indicate the possibility of Verbal-Performance IQ discrepancies with the WAIS in comparison to the Wechsler-Bellevue. The one study reporting both tests on the same 50 Ss , drawn from a university counseling center, indicates that WAIS Performance IQs averaged 4.52 points lower than the Verbal Scale IQs. However, no difference existed between the WB Verbal and Performance IQs. A second study found Performance IQs significantly lower (5.69 IQ points) than Verbal IQs in 154 psychiatric patients, but again found no difference between the WB Verbal and Performance IQs in 392 comparable patients. Similar Verbal-Performance IQ discrepancies are reported for a sample of 130 normal, aged Ss on the WAIS (Eisdorfer, Busse, & Cohen, 1959).

The average Performance minus Verbal IQ difference in the epileptic groups in this present study is -4.65 , almost the same as that noted above. To check the possibility that our epileptic Ss may not differ significantly from the general population of the hospital, the WAIS records of 100 patients under 60

TABLE 7

MEAN AGE, EDUCATION, PERFORMANCE, AND VERBAL IQs OF 100 MALE HOSPITALIZED SUBJECTS WITHOUT DIAGNOSIS OF ORGANIC BRAIN DAMAGE

	Age	Educa- tion	Perform- ance IQ	Verbal IQ	Perform- ance- Verbal
\bar{X}	33.98	9.01	86.69	90.19	-3.50^*

* Significant at the .01 level.

years of age who had no diagnosis of brain damage, and who had not been used in present study were selected at random from the Durham VA Hospital files. As indicated in Table 7, this group is comparable in age and education to the groups in the present study. The results are clear: the Performance IQ is 3.50 points below the Verbal IQ, a significant difference ($p < .01$). The Performance-Verbal discrepancies of the two groups of epileptics do not differ significantly from this larger group.

From these results it is concluded that our epileptic groups do not manifest the previously reported preoperative Performance-Verbal IQ discrepancies. Further, should other investigators using different populations substantiate the lower Performance IQ relative to Verbal IQ reported in the four studies cited above, the implication for the clinician is obvious: some readjustment in the concept of the magnitude of the discrepancy necessary before differences in those abilities represented by these scales can be inferred.

Confirmation of Prediction 2 in the Generalized group vs. Controls suggests that pattern of subtest performance may reflect the effects of cortical dysfunction of an epileptic nature in patients who may not be manifesting behavioral signs of epilepsy at the time of intelligence testing. It is of interest to note that the two subtests which discriminated best, Similarities and Picture Arrangement, have been identified by previous investigators as susceptible to temporal lobe malfunction (Milner, 1954; Reitan, 1957).

Further, the effect of degree of abnormal activity is noteworthy (Tables 5 and 6). The Moderate-Marked Activity group had a much lower Performance IQ than Verbal while the Minimal Activity group had practically identi-

cal Verbal and Performance IQs. The presence of subclinical neurophysiological disturbances may have a more telling effect on the timed tests constituting the Performance Scale. Indeed, *all* the Performance subtests were lower for the Moderate-Marked group than the Minimal Activity group despite the former's educational advantage. Such results are in accord with the impaired attentiveness and deficits in sustained serial activity reported in similar groups by Hovey and Kooi (1955); Morrell (1956); and Rosvold, Mirsky, Sarason, Bransome, and Beck (1956).

SUMMARY

Patterns of cognitive functioning on the WAIS were studied in three groups of Ss. Fifteen psychomotor (temporal lobe) epileptics with unequivocal unilateral EEG foci; 16 generalized seizure patients with temporal lobe foci; and 15 nonCNS damaged Controls. Ss were matched on relevant variables and were similar in degree and pattern of personality disturbance (MMPI). Contrary to previous reports, the Psychomotor group does not manifest a deficit in Verbal IQ relative to Performance IQ. The Generalized group is significantly lower, as predicted, on a combined Information, Similarities, Digit Span, and Picture Arrangement Score. Significantly lower Performance IQs are found in a moderate and very active abnormal EEG activity group compared to a minimal activity group, regardless of seizure pattern.

From data based on 100 hospitalized Ss in this study and studies cited by Wechsler, it is suggested that, unlike the Wechsler-Bellevue, the Performance IQ on the WAIS tends to be three to five points lower than the Verbal IQ.

REFERENCES

- EDWARDS, A. *Experimental design in psychological research*. New York: Rinehart, 1950.
- EISDORFER, C., BUSSE, E. W., & COHEN, L. D. The WAIS performance of an aged sample: The relationship between verbal and performance IQs. *J. Geront.*, 1959, 14, 197-201.
- GIBBS, F. A., & GIBBS, ERNA L. *Atlas of electroencephalography*. Vol. II. *Epilepsy*. Cambridge, Mass.: Addison-Wesley, 1952.
- HOVEY, J. B., & KOOI, K. A. Transient disturbances of thought processes and epilepsy. *AMA Arch. Neurol. Psychiat.*, 1955, 74, 287-291.
- MEYER, V. H., & JONES, H. G. Patterns of cognitive test performance as functions of the lateral localization of cerebral abnormalities in the temporal lobes. *J. ment. Sci.*, 1957, 103, 758-772.
- MILNER, BRENDA. Intellectual functions of the temporal lobes. *Psychol. Bull.*, 1954, 51, 42-62.
- MILNER, BRENDA. Psychological defects with temporal lobe excision. In H. C. Solomon, S. Cobb, & W. Penfield (Eds.), *The brain and human behavior*. Baltimore: Williams & Wilkins, 1958. Pp. 244-257.
- MORRELL, F. Interseizure disturbances in focal epilepsy. *Neurology*, 1956, 6, 327-333.
- PENFIELD, W., & JASPER, H. *Epilepsy and the functional anatomy of the human brain*. Boston: Little, Brown, 1954.
- QUADFASSEL, ANGELA F., & PRUYSER, P. Cognitive deficit in patients with psychomotor epilepsy. *Epilepsia*, 1955, 4, 80-90.
- REITAN, R. Psychological effects of cerebral lesions in human beings. Paper read at the Midwestern Psychological Association, Chicago, May 1957.
- ROSVOLD, H. E., MIRSKY, A. F., SARASON, I., BRANSOME, E. D., JR., & BECK, L. H. A continuous test of brain damage. *J. consult. Psychol.*, 1956, 20, 343-350.
- WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corp., 1955.
- WECHSLER, D. *The measurement and appraisal of adult intelligence*. Baltimore: Williams & Wilkins, 1958.

(Received September 17, 1959)

ON THE BREAKDOWN OF THE SENSE OF REALITY:

A COMMENT

PHILIP A. GOLDBERG

University of Buffalo

Hozier (1959) has attempted to answer a difficult question. Accepting the breakdown of reality as the central fact of schizophrenia, she hypothesizes that, "the loss of the sense of reality in the schizophrenic individual involves a breakdown in the bodily self as a consequence of insufficient cathexis" (p. 186). As a test of this hypothesis Hozier investigated the spatial perception of schizophrenics by means of a Figure Placement task, a Doll task, and a Draw-a-Person task. Her results, which she accepts as confirming her hypothesis, show that schizophrenics make significantly more errors on these tasks than normal controls.

Unfortunately for the clarity of her results there are methodological and theoretical difficulties which she has either overlooked or too cavalierly dismissed. There has been much work reported in the literature to indicate that the generally poor performance of schizophrenics can, in large part, be attributed to motivational variables. Cavanaugh (1958) found this to be the case with concept formation tasks, and Lang (1959) obtained similar results investigating reaction time. Lang (1959) believes that "Shakow's concept [1950] of an inability to maintain set appears to fit the data" (p. 267). Hozier mentions Shakow's concept but dismisses it on the grounds that her experimental group consisted of patients judged to be cooperative by their ward psychiatrist.

One has no way of knowing or evaluating the adequacy of the ward psychiatrists' ratings of cooperativeness. Hozier does not tell us the procedures used to obtain these ratings, nor does she specify the variables that entered into the determination. One might speculate that the demands of custodial care might lead ward psychiatrists to equate tractableness with cooperativeness. Without pursuing this speculation, it would seem fair to

say that when discussing schizophrenic performance, problems of motivation and set are too important for the reader to be confident that Hozier's solution is adequate.

The literature demonstrating schizophrenic deficit is too voluminous to need citation here. Hozier did not merely attempt to indicate another area of deficit; her attempt was to demonstrate the relationship between reality loss and "the spatial problem of dealing with the body . . . and the relationship of the body to the world" (p. 194). But how would her schizophrenics perform in other task situations that do not specifically involve such spatial perception of the body? It is possible, perhaps likely, that they would show deficit. But whatever the case, Hozier's study does not have this control and the reader must question the correctness of her conclusion, which relates the schizophrenics' impaired performance to the specific problem of the spatial perception of the bodily self. Hozier has arbitrarily explored a single facet of a multifaceted problem to emphasize bodily self as the primary model of reality. A more parsimonious explanation of her data might be formulated in terms of attitudinal and attentional variables presumed to operate in a wide range of task situations.

Hozier does raise tangentially this question of confounding when she states, "what part other variables may play in accounting for the significant results cannot be determined by the present study" (p. 193). But here it seems her primary concern is with the adequacy of her task situations. She dismisses the question even before she raises it when she says, "certainly, the tasks are related to spatial relations and organization" (p. 193). Accepting on faith the adequacy of her tasks, Hozier's recognition that other variables might play a role in her results emphasizes the point made above. It is not

necessary to question that some relationship exists between her tasks and spatial organization; it is necessary to question the control of the "other variables" that may also be related to the tasks and which might have made Hozier's results significant. "Other variables" always exist potentially, and no one has the right to demand of an experimenter that he exhaustively anticipate these variables. However, when the accumulated research of other workers implicates particular classes of variables as being significantly related to a phenomenon, it is the responsibility of the experimenter investigating this phenomenon to control and account for these variables.

There are other questions that might be raised about this study. Hozier states that patients who were receiving insulin coma or electric shock treatment were excluded from her experimental group. Presumably this was done to insure that her subjects' (Ss') performance would not be hampered by extraneous factors. But it is widely recognized that one of the major obstacles in doing research with hospital patients is the difficulty in getting patients who are not "on drugs." Hozier does not report whether the schizophrenics in her experimental group were receiving drugs. If they were, we can only speculate as to the probable effect on performance.

Another question might be raised about Hozier's use of one-tailed tests. Goldfried (1959) would question her criterion for the selection of a one-tailed test, carrying with it the possible increase in the level of significance. Nevertheless, this issue is controversial and though this writer would agree with Goldfried, it is perhaps not a crucial criticism. However, other criticisms already made do seem to be crucial.

But it would be unfair and scientifically unprofitable to ignore the value inherent in Hozier's work because of what seem to be certain significant weaknesses. Those features believed to be methodologically weak are not intrinsic to the theoretical problem at issue. Had the Ss been asked to draw three dimensional geometric figures in addition to drawing a person, the study might have been tightened considerably. Such a task would presumably not involve "the spatial problem of dealing with the body," but it would ap-

pear to tap essentially the same motoric and spatial skills. Thus, it might provide one needed control of set variables. Certainly this is just one possible way, and undoubtedly not the best, to make Hozier's study a more rigorous investigation.

Hozier has proceeded with an interesting theoretical approach to attack imaginatively a difficult problem. This is not meant to be the traditional sop with which critics conclude their papers. Hozier's study is worth criticizing because of its fundamental value. Should a more tightly designed study along essentially the same lines as Hozier's work obtain similar findings, its implication for ego psychology and for an understanding of schizophrenic processes would be of undeniable worth. Additionally, such a study might help to resolve the controversy over the nature of schizophrenic deficit. It has been suggested that schizophrenic deficit represents a withdrawal response to threatening stimuli (Rodnick & Garmezy, 1957). Cavanaugh (1958), Lang (1959), Hunt and Cofer (1944), and others have stressed the motivational variables involved in schizophrenic deficit. In short, we have yet to understand why the schizophrenic behaves as he does. Hozier's work expands a new approach to an old question. We need new approaches.

REFERENCES

- CAVANAUGH, D. Improvement in the performance of schizophrenics on concept formation tasks as a function of motivational change. *J. abnorm. soc. Psychol.*, 1958, 57, 8-12.
- GOLDFRIED, M. R. One-tailed tests and "unexpected" results. *Psychol. Rev.*, 1959, 66, 79-80.
- HOZIER, ANN. On the breakdown of the sense of reality: A study of spatial perception in schizophrenia. *J. consult. Psychol.*, 1959, 23, 185-194.
- HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 971-1032.
- LANG, P. J. The effect of aversive stimuli on reaction time in schizophrenia. *J. abnorm. soc. Psychol.*, 1959, 59, 263-268.
- RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. Jones (Ed.), *Nebraska symposium on motivation*. Vol. 5. Lincoln: Univer. Nebraska Press, 1957. Pp. 109-184.
- SHAKOW, D. Some psychological features of schizophrenia. In M. L. Reymert (Ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950. Pp. 383-390.

(Received September 30, 1959)

A REPLY TO GOLDBERG

ANN HOZIER

Veterans Administration Mental Hygiene Clinic, Los Angeles

Goldberg, after a thoughtful and careful review of my article, raises some cogent criticisms. His main and perhaps his most crucial criticism is of my failure to control for motivational variables. In support of this criticism, he cites experimental work which concludes that the poor performance of the schizophrenic can be largely accounted for by motivational variables. Attractive as this explanation may be, I see it as being tautological, in that by definition we characterize schizophrenics as being motivationally impaired. They are apathetic, withdrawn, disinterested, distractible, lacking in goal direction, etc. If we attempt to account for their behavior in terms of motivational variables, we are only attempting to affirm our definition. Such an interpretation is qualitatively similar to saying that a mental defective's lack of intelligence hampers his performance. Perhaps the kind of question we should ask is what underlies the impairment of the ability to maintain set and attend? Do motivational variables constitute the necessary and sufficient conditions for developing a stable sense of reality? Our hypothesis is that the ability to attend, to direct one's energies to a task, to make adequate use of stimuli, to think abstractly, etc., is dependent upon an adequate differentiation of body/world or, in our terms, on a stable, well-cathexed bodily self. As long as this differentiation is not clearly delineated, then true subject-object relations cannot exist on a stable consistent level. Motivated behavior, in its true meaning, is predicated on the capacity of the subject (*S*) to relate, to direct himself to an object. We assume that the schizophrenic is incapable of sustained motivated behavior. That is, he does not experience himself as an *S* in relationship to an object with any consistency as a consequence of insufficient cathexis of his

bodily boundaries. He does have affective cathexis available to him, but it is attached to various and sundry, isolated things.

I assumed that the schizophrenics would have no vested interest in the tasks—that they would not be motivated because of the very nature of their illness. The word cooperative, rather than motivated, was used to indicate that the interest of the schizophrenic was solicited rather than spontaneously given. I further assumed that cooperation was not a global trait. We all have observed that the schizophrenic, like most people, will be cooperative with one person and not with another or will do one task and not another at request. The psychiatrists were asked to select patients who met the selection criteria (Hozier, 1959, p. 186) and whom they judged would be cooperative with the investigator in doing the particular tasks. The psychiatrists were aware of the nature of the tasks, but were not aware of the predictions. They selected 36 patients, some of whom required maximum supervision because of their destructive and assaultive behavior. Other patients resided on an open or semiopen ward. Some of these patients were either already working in the community or in the process of obtaining positions outside of the hospital. The records of the 36 patients were reviewed, and it was found that 3 of them did not meet one of the selection criteria. The patients were seen in the order of their appearance on the list. Of the 26 *Ss* contacted, the cooperation of 1 *S* could not be obtained.

Are clinical evaluations of cooperativeness adequate? Did the investigator fail to see evidences of uncooperativeness due to her own involvement in the research? Unfortunately, an unequivocal answer to these questions cannot be given. It is the investigator's impression that the schizophrenics' failures or

inadequacies in dealing with the tasks were not largely a consequence of their unwillingness to attempt to deal with them.

Goldberg raises an important question when he asks how would schizophrenics perform on tasks not specifically involved in the spatial perception of the body. Theoretically, as long as the body is not sufficiently cathected to become bounded and differentiated from everything that is not the body, there exists no frame of reference from which to judge the reality of events in the external world or the reality of one's own psychological experiences. The conventional dimensions of space are body dimensions. What is called up or down, to the left or right, or to the front or behind does not depend on the position of the object, but on the position of the body in relationship to the object. Thus, we would hypothesize that if the boundaries of the body are not sufficiently cathected, then there will be a disturbance in spatial perception of non-human as well as human objects. We have seen this, for example, in the schizophrenic's perception of the vertical (Carini, 1955), in his handling of the Bender-Gestalt (Bender, 1938), in his dealing with the Body Space Test (Zierer, 1950), etc.

I would expect if he were given, for example, a formal test of spatial perception or geometry problems in which he can depend on his knowledge of rules, propositions, and theorems of handling space that he would have significantly less difficulty than if the task required that he rely on his experiencing his relationship to and in space. Perhaps it will make it clearer to illustrate the distinction between reality testing and a sense of reality. Depersonalization, estrangement, and hypochondriasis, often early signs of schizophrenia, exemplify this difference. The person may *experience* himself as changed, experience his world as changed, or experience himself as physically sick, yet at the same time, *know* that he is the same person, that the world hasn't changed, and that his physical health is unimpaired. Both reality testing and a sense of reality are necessary for dealing with the world, but in general, reality testing refers to testing something via the intellectual processes, and a sense of reality, of experiencing something affectively. The more the

task can be adequately dealt with by a pure knowledgeable approach, the less, it is hypothesized, will the schizophrenic fail. Had I used the type of control mentioned above, it undoubtedly would have tightened the research considerably.

Goldberg is correct that I was concerned with the adequacy of the tasks. When he says, "It is not necessary to question that some relationship exists between her tasks and spatial organization; it is necessary to question the control of the 'other variables' that may also be related to the tasks and which might have made Hozier's results significant," I believe his concern here is again with the motivational variables. I have already attempted to indicate why I am not concerned about these variables in the same way as he.

Recent ECT and Insulin Coma Therapy were controlled not because they were extraneous factors *per se*, but because of the confusion which so often accompanies their use. Confusion has not been noted as a consequent of tranquilizers unless used in toxic dosages. In those cases, the patients would have had neurological signs and on that basis have been excluded. The confusion the patients had was part and parcel of their illness. Some of the patients were on drugs and some not; however, the differences in their performances were not analyzed. I would predict that of the patients having an equal degree of disturbance those on tranquilizers would perform more adequately than those not on them.

As Goldberg indicates, there has been considerable controversy about the use of the one-tailed test. As to the criteria for its use, I continue to agree with Jones (1954, p. 586).

Goldberg has raised some significant and crucial questions about my research. Although I disagree with him on his major point of the control of motivational variables, I would agree that some question may be raised about the adequacy of clinical judgments of cooperativeness. However, as I have thought of this criticism, I must admit that I find it difficult to conceive of an adequate way to control this variable which at some level or other does not lean heavily on clinical judgment. His point about the use of a control task is well made. I would expect the use of

drugs to exert its effect in the direction of nullifying the hypotheses. Thus, I doubt that the lack of control of drugs contributes to the significant results.

There is considerable need for further research not only to cross-validate my findings with some of the controls mentioned above, but to investigate other hypotheses derived from the theory. It is impossible for me to state the importance of the bodily self in a clearer way than Scott (1951), when he says,

The body image is not just an object of study in the sense that science has made so many things objects of study. It is essentially something that one lives through and in. To-day when there is so much disembodied knowledge around, and so many psychological forces loose in the world belonging to no one, and when we are in very real danger of being devoured by our own creations, it seems significant that the human image in the midst of all

of these should be brought to the forefront, for it is in the last resort the only mediator of them (p. 266).

REFERENCES

- BENDER, LAURETTA. *A visual motor gestalt test and its clinical use*. New York: Amer. Orthopsychiat. Ass., 1938.
- CARINI, L. *An experimental investigation of perceptual behavior in schizophrenia*. Unpublished doctoral dissertation, Clark University, 1955.
- HOZIER, ANN. On the breakdown of the sense of reality: A study of spatial perception in schizophrenia. *J. consult. Psychol.*, 1959, 23, 185-194.
- JONES, L. A rejoinder on one-tailed tests. *Psychol. Bull.*, 1954, 51, 585-586.
- SCOTT, R. The psychology of the body image. *Brit. J. Psychol.*, 1951, 24, 254-256.
- ZIERER, E. Dynamics and diagnostic value of creative therapy: I. The body space test. *Acta med. Orient., Jerusalem*, 1950, 9, 35-42.

(Received December 14, 1959)

SEX-ROLE PREFERENCES IN THREE- AND FOUR-YEAR-OLD CHILDREN

WILLARD W. HARTUP AND ELSIE A. ZOOK

Iowa Child Welfare Research Station

The rubric of sex-role differentiation has been used to cover both the processes leading to sex differences in overt behavior (e.g., boys are more physically aggressive than girls) and the introjection of values and attitudes that are more appropriate in a given culture for one sex than the other. A recent attempt at clarifying the concept of sex-role differentiation has been made by Brown (1956, 1957) with the assistance of Robert Sears, and Lynn (1959). These authors have distinguished among *sex-role identification*, *sex-role preference*, and *sex-role adoption*. The development of masculinity and femininity is thus defined in terms of at least three logically separated psychological processes. The present study was concerned only with sex-role preference which has been defined by Brown (1956) as "behavior associated with one sex or the other that the individual would like to adopt or that he perceives as the preferred or more desirable behavior."

The sex-role preferences of children have been measured in two ways. Rabban (1950) developed a technique which consisted of asking children to state their preferences from a group of 16 toys, 8 commonly associated with girls and 8 associated with boys. The administration of this test to 300 children from three through eight years of age showed that: (a) boys possess clear-cut preference patterns at an earlier age than girls; and (b) strong appropriate preference patterns appear earlier among lower-class children than among middle-class children.

Brown (1956) measured children's sex-role preferences by means of a projective test called the *It Scale for Children*. In taking this test the child chooses between pictures of various objects commonly associated with one

sex or the other (e.g., toys, clothes, household objects, games, etc.). The choices are not made for the child himself but for "It," a drawing of a sexless figure. Brown has reported normative data for 146 kindergarten children from Denver, Colorado (1956) and 613 children from kindergarten through fifth grade from Pleasanton, California (1957). These data have shown that: (a) distinctive sex-role preferences existed for boys and for girls at all ages studied; (b) kindergarten boys were masculine in their preferences but older boys were even more masculine in their preference scores; (c) kindergarten girls had "mixed" preferences and older girls slightly *masculine* preferences; (d) at all age levels girls' preference scores were more variable than boys' scores.

The present study was designed to extend Brown's work with the It Scale for Children. Preferences of the kind measured by the It Scale are undoubtedly acquired before the child is of kindergarten age. One purpose of this investigation was to obtain information concerning the sex-role preferences of three- and four-year-old children.

A second purpose of the present study was to explore the effects of the instructions employed with the It Scale. It has been mentioned previously that the subject's (S's) choices are made for an ambiguous drawing of a child rather than for S himself. The similarity of the pictured figure to S has been shown to be a relevant variable to children's performance on some projective tests (Armstrong, 1954; Bills, 1950; Furuya, 1957) but irrelevant to performance on others (Biersdorf & Marcuse, 1953). If "stimulus-to-subject similarity" should affect performance on the It Scale, it is obvious that any one pro-

When mean scores from the two administrations of the It Scale were compared it was found that girls moved toward greater femininity on the retest (see Table 1). It is difficult to account for this change in terms of inter-S communication, practice effects, etc., considering the ages of the Ss and the length of the interval between tests. It is possible that the test-retest differences reflect new attitudes acquired by girls during the two-month interval between tests. Interestingly, boys did not move toward greater masculinity on the retest. This would imply that the preschool period may not be so important for the male as for the female as a transitional period in sex-typing.

RESULTS

The It Scale scores of 150 Ss were analyzed by means of an analysis of variance that involved three factors: age, sex, and instructional procedure. The frequency distributions of the 12 subgroups appeared by inspection to approximate normality. The variances of these subgroups appeared to be homogeneous with two exceptions: (a) the variance for four-year-old girls in Group A was approximately four times the average variance of the other subgroups; (b) the variance for four-year-old boys in Group C was approximately one-third the average variance of the other subgroups (see Table 2). While two variances thus appeared to be deviant, the possibility remained that homogeneity might be true of the table as a whole due to the apparent homogeneity of variance in 10 of the cells. Accordingly, Bartlett's test was applied which yielded a chi square of 17.86, $df = 11$, $p < .10$. Thus the hypothesis of homogeneity of variance could be sustained. It was decided, however, that no findings from the

TABLE 2

MEANS, RANGES, AND STANDARD DEVIATIONS OF SCORES ON THE IT SCALE FOR BOYS AND GIRLS IN TWO AGE GROUPS AND THREE INSTRUCTION GROUPS

	Instruction Group	N	Mean	Range	SD
Girls					
Three-year-olds	A	10	37.70	20-49	10.071
	B	10	31.10	0-50	15.667
	C	10	32.30	15-47	12.475
Four-year-olds	A	15	31.27	0-84	25.058
	B	15	11.67	0-40	11.948
	C	15	15.80	0-50	14.386
Boys					
Three-year-olds	A	10	56.50	37-74	11.342
	B	10	56.00	32-74	14.156
	C	10	62.90	35-76	11.344
Four-year olds	A	15	60.80	30-84	13.668
	B	15	63.33	31-84	17.117
	C	15	71.13	48-84	8.155

analysis of variance would be accepted as significant unless the probability level reached was beyond .01. This decision was based on recommendations by Lindquist (1953).

The results of the analysis of variance may be found in Table 3. In this analysis the effects of sex were highly significant ($F = 220.762$, $p < .001$). While on the one hand this finding merely demonstrates the validity of the It Scale, it is also important because it demonstrates clear-cut sex-role differentiation among very young children.

The age by sex interaction was also significant ($F = 16.464$, $p < .001$) and may be interpreted as meaning that the effects of age on It Scale scores depended upon the sex of the child. The simple effects of age were studied by means of t tests in which the error term from the analysis of variance was employed. The age difference for girls was significant ($t = 3.925$, $p < .001$), indicating that four-year-old girls were more feminine in their sex-role preferences than three-year-old girls. The age difference for boys, on the other hand, reached only borderline significance ($t = 1.840$, $p < .10$). Table 2 shows that the direction of this trend was for four-year-old boys to display more masculine preferences than three-year-old boys.

The effects of the instructions on It Scale scores also depended on the sex of the child. The instruction by sex interaction was significant beyond the .005 level. Analysis of the

TABLE 1

MEAN IT SCALE SCORES ON TEST AND RETEST

	Mean	<i>SD</i>	<i>t</i>	<i>p</i>
Girls (<i>N</i> = 20)				
Test	34.60	14.32	2.675	<.02
Retest	25.85	16.34		
Boys (<i>N</i> = 20)				
Test	60.30	15.11	.974	—
Retest	63.33	16.93		

TABLE 3
ANALYSIS OF VARIANCE OF IT SCALE SCORES

Source	df	Sum of Squares	Mean Square	F
Instructions	2	1,193.45	596.73	2.537
Age	1	506.25	506.25	2.152
Sex	1	51,931.20	51,931.20	220.762**
Instructions \times Age	2	151.64	75.82	—
Instructions \times Sex	2	2,813.46	1,406.73	5.980*
Age \times Sex	1	3,872.99	3,872.99	16.464**
Instructions \times Age \times Sex	2	457.21	228.61	—
W cells	138	32,462.63	235.236	
Total	149	93,388.83		

* $p < .005$.

** $p < .001$.

simple effects showed: (a) girls in Group A showed less feminine preferences than girls in both Group B ("little girl"; $t = 3.320$, $p < .01$) and Group C (S's name; $t = 2.637$, $p < .01$); (b) mean scores for girls in Groups B and C did not differ; (c) boys in Group A did not differ from boys in Group B, but were less masculine than boys in Group C ($t = 2.019$, $p < .05$); (d) boys in Group B were slightly less masculine than those in Group C ($t = 1.715$, $p < .10$).

Inspection of the raw data relative to the analysis of variance showed confounding of both birth order and size of family with the variable of age. A greater proportion of three-year-old Ss as compared with four-year-olds were first born children ($p < .10$). Also, four-year-old boys had more siblings than three-year-olds ($p < .02$). The same tended to be true of girls, although only at the .20 level. The results of a series of t tests, applied to the age differences with both birth order and family size held constant, were consistent with the analysis of variance results. Four-year-old girls were more feminine than three-year-olds ($p < .01$); four-year-old boys were more masculine than three-year-olds, but only at a borderline level ($p < .10$).

The analysis just reported yielded one other finding of some importance. There were no significant differences between It Scale scores of first born and later born children. Thus birth order, per se, appeared to have no relation to the early acquisition of sex-role preferences.

Although this study was not designed to

relate sex-role preferences to socioeconomic status, Ss were obtained from two diverse social classes. However, 61% of the lower-class children (36 Ss from the community-supported day care center) had fathers not living at home. When compared with the other Ss tested, who were from intact middle-class homes, no class differences were found in sex-role preference scores for either boys or girls. For heuristic purposes, the 14 lower-class Ss who came from intact homes were compared with the middle-class Ss. Once again, no class differences in It Scale scores emerged.

DISCUSSION

Sex Differences

The sex difference found in It Scale scores suggests that at least some aspects of sex-role differentiation begin very early in life. Beginning in infancy, parents in the United States commonly prescribe different toys, modes of dress, etc., for boys and for girls. The findings of the present study reflect the success of such child training procedures.

A sex difference in degree of appropriate sex-role preferences was also revealed by the data. In middle childhood, boys prefer the stereotyped masculine role more strongly than girls prefer the stereotyped feminine role (Brown, 1957). The same appeared to be true of the three-year-olds in the present study (see Table 2). Among four-year-olds in Groups A and C boys also showed slightly stronger masculine preferences than girls showed feminine ones. Only in Group B were four-year-old girls more feminine than boys

were masculine. Thus, the majority of these data are in line with the findings for older children: boys more strongly prefer the stereotyped male role than girls prefer the stereotyped female role. It should be stressed, however, that this tendency was considerably less marked among the preschool Ss of the present study than among the older Ss of other studies. Quite probably the preschooler is only beginning to experience: (a) the greater socio-cultural advantages and more consistent rewards for sex-typed behavior accorded the male in United States culture; (b) the more clearly defined sex-role for the male than for the female; and (c) more frequent punishment of boys than girls for opposite-sex behavior.

Age Differences

Although only significant at a borderline level, four-year-old boys obtained more masculine scores on the It Scale than three-year-old boys. A similar relation between age and sex-role preferences has been reported for older boys (Brown, 1957). In the present study three-year-old boys obtained a mean score of 58.47; the 11-year-old boys tested by Brown obtained a mean score of 76.73. Thus, for boys, there appears to be a steady change toward greater masculinity during the 3-through 11-year period. This change suggests that boys probably receive relatively consistent reinforcement throughout early and middle childhood for adopting certain stereotyped aspects of the male role. This finding could also reflect a condition wherein the early parental demands for appropriate male sex-typing are augmented in later years by the demands of other socializing agents (e.g., teachers, peers, etc.).

The present findings showed that four-year-old girls made significantly more feminine choices than three-year-old girls. By itself, this finding is analogous to the age difference found for boys. However, Brown (1957) reported a change toward *masculinity* in older girls: kindergartners obtained a mean score of 38.40 while fourth-grade girls reached 56.40. In the present study the three- and four-year-old girls in Group A (whose scores are most comparable to Brown's) obtained mean scores of 37.70 and 31.27, respectively.

Therefore, preschool-aged girls possess the most feminine sex-role preferences of any group of girls in the 3- through 10-year period. Actually, the age differences found in It Scale scores for girls support an hypothesis of Lynn's (1959): girls may be femininely oriented in early childhood due to a basic identification with their mothers, but later exposure to the masculine-oriented culture breaks down or covers over this identification during the years of middle childhood. It could be also that the mother, who is virtually the sole feminine model for the preschool girl, is a model that is particularly stereotyped (i.e., the mother's femaleness is typified for the child by homemaking, child care, etc.). On the other hand, the older girl has additional feminine models available to her (outside the home, on television, in fiction, etc.). These models are frequently observed in nonstereotyped situations or in situations where their behavior is more typical of men than women. Such a broadened conception of the female role for the older girl could produce more masculine scores on the It Scale in which femininity is defined in terms of highly stereotyped choices.

Effects of Instructions

Instructions that employed S's name in association with the drawing resulted in more masculine scores for boys than instructions that stressed lesser degrees of stimulus-to-subject similarity. It is possible that high stimulus-to-subject similarity elicited the most valid measure of boys' sex-role preferences since the questioning was related to S himself. On the other hand, the use of S's name could have aroused anxiety concerning E's approval of inappropriate (i.e., female) preferences. Such anxiety could have elicited an inordinate number of safe, stereotyped male preferences. Thus the data indicate only that the instructions employed with the It Scale affect that sex-role preference scores of boys; the data do not indicate which instructions produce the most valid scores.

The finding that no difference existed between scores for boys in Group A ("It") and Group B ("little boy") can be interpreted in two ways. First, boys in Group A may have

perceived the figure as a boy.³ If such were the case, the test situation for Group A would be very similar to the situation for Group B. The alternative interpretation is that Ss in Group B were indifferent to the label "boy." This is saying in effect that Ss in neither group perceived the figure as a boy, a situation which seems improbable. Since the findings demonstrate that stimulus-to-subject similarity affects boys' responses on the It Scale, the various *post hoc* hypotheses concerning the dynamics of these effects become important problems for future research.

The instructions employed with the It Scale also influenced the scores of girls. For all girls, but particularly for four-year-olds, instructions that referred to the figure as "a little girl" resulted in more feminine scores than the label "It." The use of S's name produced little additional femininity in girls' scores. Since labeling the figure "a little girl" produced such marked femininity in girls' scores, the question is raised as to whether some girls in Group A perceived "It" as a boy. The hypothesis that the sexless figure was perceived as a boy by some Ss is supported by the findings that four-year-old girls in Group A had much more variable scores than girls in Groups B and C. However, variability among Group A three-year-olds was not greater than among three-year-olds in Groups B and C. If some girls perceived "It" as a boy why, then, did three-year-olds make so few masculine choices? Possibly the three-year-old girls did not possess adequate information concerning the components of masculine behavior. If so, Ss of this age who perceived "It" as a boy would have been unable to give consistently masculine choices. On the other hand, the four-year-old girls may have been sufficiently sophisticated concerning "boyishness" that they were able to choose appropriately for the figure if

they saw "It" as a boy. This interpretation would explain the fact that the differences in both means and in variability were larger among the subgroups of four-year-old girls than among the subgroups of three-year-old girls.

These findings carry particularly strong implications for the use of the It Scale with girls. If the unstructured drawing is frequently seen as a male, then older girls' responses under this condition obviously do not indicate the extent to which they prefer the female role.

Birth Order

The data from this study indicated the irrelevance of birth order to the acquisition of sex-role preferences. It should be pointed out that it is not known if older siblings of a particular sex (e.g., only older brothers or older sisters) affect the development of such preferences. The present investigation did not offer a means of checking findings of this kind that have been reported by Koch (1956) and Brown (1956).

Social Class

The fact that mean scores for the two socioeconomic groups did not differ in the present study could possibly be due to the particular lower-class population studied. One would predict on the basis of Rabban's (1950) findings that scores for lower-class children would be more sex-typed than scores for middle-class children. On the other hand, Pauline Sears (1951), in a study of doll play aggression, found that boys whose fathers were absent from the home were less aggressive (and as such, perhaps more feminine) than boys whose fathers were present. Since 61% of the lower-class Ss in this study had fathers who were absent, it is possible that the effects of social class and father absence canceled each other out in the comparison between lower- and middle-class Ss. No differences were found between the lower class Ss whose fathers were present and the middle-class Ss. However, the probability of a gross sampling error in this comparison is extremely high in view of the fact that less than 10 children of each sex made up the lower-class sample.

³ Subsequent to collecting the data for this investigation, the authors polled 10 male Ss in another nursery school group concerning what was pictured on the card. Six of these Ss said it was a picture of a girl. Obviously, data for many more Ss are needed before evaluating the validity of the hypothesis that the figure is a definitely "boyish" figure for preschool aged boys. The results of the poll are mentioned only to suggest that this hypothesis may not be tenable.

SUMMARY

The purposes of this study were: (a) to acquire normative data relative to sex-role preferences in three- and four-year-old children, and (b) to study the effects of verbal instructions that stressed varying amounts of similarity between a drawing and S himself on a projective test of children's sex-role preferences.

Ss were 161 three- and four-year-old children drawn from four nursery school populations in Iowa. Sixty-five Ss, ranging in age from 3-0 to 4-0, constituted the three-year-old group and 96 Ss, ranging from 4-0 to 5-0, constituted the four-year-old group.

Sex-role preferences in this study were measured by the It Scale for Children. The major findings: (a) clear-cut sex differences in It Scale scores were found; (b) girls at four years scored significantly more feminine than three-year-old girls; (c) four-year-old boys were more masculine than three-year-old boys at a borderline level of significance; (d) girls responded with more feminine scores when the drawing employed in the It Scale was called "a little girl" than when called "It"; (e) boys responded with more masculine scores when the figure was called by S's own name than when the figure was called "It."

While these findings imply that early childhood is an important period in sex-role development, they also imply that the acquisition of sex-role preferences by the male is a less complicated developmental process than

for the female. The findings also suggest that the measure of sex-role preferences provided by the It Scale is highly sensitive to variation in the instructions given to the S.

REFERENCES

- ARMSTRONG, MARY A. S. Children's responses to animal and human figures in thematic pictures. *J. consult. Psychol.*, 1954, 18, 67-70.
- BIERSDORF, KATHRYN R., & MARCUSE, F. L. Responses of children to human and animal pictures. *J. proj. Tech.*, 1953, 17, 455-459.
- BILLS, R. E. Animal pictures for obtaining children's projections. *J. clin. Psychol.*, 1950, 6, 291-293.
- BROWN, D. G. Sex-role preferences in young children. *Psychol. Monogr.*, 1956, 70(14, Whole No. 287).
- BROWN, D. G. Masculinity-femininity development in children. *J. consult. Psychol.*, 1957, 21, 197-202.
- FURUYA, K. Responses of school children to human and animal pictures. *J. proj. Tech.*, 1957, 21, 248-252.
- KOCH, HELEN L. Attitudes of young children toward their peers as related to certain characteristics of their siblings. *Psychol. Monogr.*, 1956, 70(19, Whole No. 426).
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- LYNN, D. B. A note on sex differences in the development of masculine and feminine identification. *Psychol. Rev.*, 1959, 66, 126-135.
- RABBAN, M. Sex-role identification in young children in two diverse social groups. *Genet. psychol. Monogr.*, 1950, 42, 81-158.
- SEARS, PAULINE S. Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, 65(6, Whole No. 323).

(Received September 30, 1959)

PERCEPTUAL-MOTOR DEVELOPMENT IN CHILDREN RETARDED IN READING ABILITY¹

FRANK M. LACHMANN

Veterans Administration Regional Office, New York

Finding causes for reading disabilities in children has intrigued the members of almost every profession having contact with such children. Although varied explanations for reading disabilities have been offered, two main approaches to the problem can be identified. Reading disability has been considered by a number of writers (Jarvis, 1958; Vorhaus, 1946) primarily as a symptom of emotional disturbances. They discussed what reading might symbolize for the child (i.e., conforming to environmental demands, a "feminine" activity for boys, the unconscious conflicts, etc.) and what the child might be rebelling against when he fails to learn to read.

A different emphasis is expressed by Bender (1949) and Fabian (1945) who view the development of reading ability as dependent upon the same factors which result in perceptual-motor maturity. Following this viewpoint, Silver (1952) includes as specific and integral parts of this syndrome functioning on the Visual Motor Gestalt Test on a rather primitive level, e.g., difficulty in constructing angles and difficulty in figure background perception. In their studies Fabian and Silver found that retarded readers show inferior gestalt test performance compared with normal readers. These results seem to confirm the hypothesis that reading disability is characterized by perceptual-motor immaturity and

hence a developmental lag. Yet these studies may be criticized for failing to assess the extent to which the emotional conflicts are primary and the developmental retardation (the reading disability) a reflection of the conflicts.

Perceptual-motor functioning, usually operationally defined by performance on the Bender gestalt test or other tests of this nature, has been described by Silver (1953) as not only involving visual perception, but also the expression of that perception, the result reflecting the quality of the perception plus the motor impulsivity and the attempts at its control. These factors are considered as inseparable. In the present study an attempt has been made to investigate this developmental framework and to control for the influence of emotional maladjustment in the retarded readers. Thus, if reading disability reflects a lag or retardation in perceptual-motor development, retarded readers should show significantly less mature performance on the Bender gestalt test than *either* a comparable group of normal readers or a group of emotionally disturbed children who have normal reading ability.

SELECTION OF SUBJECTS

Since reading disability is generally defined as an inability to learn to read properly in spite of normal intelligence, it was necessary to obtain a measure of reading ability and intelligence for each subject (*S*) to be considered for inclusion in the present study. To ascertain the reading level, the California Reading Test (Tiegs & Clark, 1950a, 1950b) was employed with reading retardation being defined as a score at or below the 20th percentile for the *S*'s chronological age and normal ability as a score at the 40th percentile or above. Normal intelligence was defined as an IQ score of 90 or more on the verbal section of the Wechsler Intelligence Scale for

¹ The present paper is a condensation of part of a doctoral dissertation completed at Northwestern University, Evanston, Illinois. The author wishes to thank Robert I. Watson for his suggestions, and Jacob Cohen of New York University for his help with the statistical section of this paper and the staffs of the Institute for Juvenile Research, Chicago; Dyslexia Memorial Institute, Chicago; Bureau of Child Study, Chicago; National College of Education, Evanston, Illinois; Michael Reese Hospital, Chicago; and Bellevue Psychiatric Hospital, New York, for their cooperation in obtaining the subjects.

TABLE 1
DESCRIPTION OF GROUPS

Variables	Reading Disability Group R		Diagnostic Groups Emotionally Disturbed Group E		Normal Group N	
	RY	RO	EY	EO	NY	NO
Number of Subjects	20	20	20	20	20	20
Ratio—Boys:Girls	19:1	14:6	19:1	15:5	19:1	14:6
Age in Months						
Mean	110	133	105	134	109	129
SD	5.7	7.5	7.0	7.1	6.5	6.4
Intelligence Level (IQ)						
Mean	104 ^a	104 ^a	111 ^b	106 ^b	107 ^a	106 ^c
SD	9.9	9.7	11.9	10.6	6.8	10.7
Reading Level in Months						
Mean	93	106	115	134	120	137
SD	7.0	12.0	9.9	9.1	6.1	11.4

^a WISC Verbal IQ.

^b WISC Verbal IQ or S-B.

^c K-A IQ.

Children (WISC) (Wechsler, 1949), or the Stanford-Binet test (S-B) (Terman & Merrill, 1937) or the Kuhlman-Anderson Intelligence Test (K-A) (Kuhlman & Anderson, 1952).²

Three groups of 40 children were selected. Group R consisted of children meeting the above criteria with respect to reading retardation and normal intelligence as measured by the WISC. They were selected from clinics for remedial reading. Children referred to mental hygiene clinics for psychological or psychiatric diagnosis or treatment who were of normal intelligence and normal reading ability comprised Group E. The reasons for which these children had been referred to the clinics may be summarized in three overlapping categories: behavior problems in school (disturbance in class, truancy, etc.); unmanageable behavior at home (stealing, temper tantrums, etc.); "nervous symptoms" (fears, enuresis, etc.). The normal group (Group N) was obtained from a public school and did not

include any children having a record of referral for diagnosis or treatment because of emotional difficulties.

Each of these groups were divided into subgroups of 20 Ss each according to age. The younger half (Subgroup Y) ranged in age from 8 to 9 years and 11 months, and the older half (Subgroup O) from 10 years to 11 years and 11 months. More detailed information concerning the groups is contained in Table 1.

For Groups E and R, if a WISC or S-B score less than one year old was available, and for Group N if a K-A score less than one year old was available, no additional IQ tests were given. Otherwise the investigator individually administered the verbal section of the WISC. The reading test was individually administered to the children in Groups R and E and in a group administration to the children in Group N.

Since reading disability is noted more frequently among boys, the groups were equated as to ratio of boys to girls, as well as age and IQ. The *t* tests computed between mean reading scores at comparable age levels for Groups E and N did not reach statistical significance. An analysis of variance computed from the IQ scores of all six groups did not yield a sig-

² Because of time limitations it was not possible to use the same intelligence test for all children. Correlation among the scores of the three tests (Pastovic & Guthrie, 1951; Traxler, 1941) for similar age groups and within the IQ range of the present study suggests sufficient comparability of scores to permit using them for equating the groups.

nificant F value. This suggests that the groups were adequately equated on the basis of reading ability and intelligence.

METHOD

The nine figures comprising the Bender gestalt test were individually administered according to the usual procedure as described by Bender (1938). Five types of distortions which, according to Bender (1938), are suggestive of immaturity in perceptual-motor development when still produced above the age of eight years, were then scored as being either present or absent for each child. Definitions of the distortions were taken from Bender (1938) and from the Bender gestalt test scoring criteria developed by Pascal and Suttell (1951).

Angulation: rounding the points at corners, blunting corners, adding extra angles, changing the direction of angles, omitting angles or leaving gaps at corners greater than one-eighth inch. Silver (1953) has related the presence of this distortion to reading disability. If two or more angles of Figures A, 3, 4, 7, or 8 showed the distortion, it was scored as present for the S .

Rotation: scored as present when in the reproduction of a figure the major axis of the stimulus figure had been turned 90 degrees. No score was given when the S turned the paper to make the most economical use of it. Fabian (1945) and Silver (1953) have related the presence of this distortion to reading disability.

Primitivation: drawing circles or loops instead of dots. This was scored as present when two dots or more of Figures 1, 3, or 5 were distorted. De Hirsch (1954) has related this and the following two distortions to reading disturbances.

Separation: separating the adjoining or overlapping parts of figures by one-eighth inch or more. It was scored present when noted in Figures A, 4, 7, or 8.

Slant: drawing the columns of circles in Figure 2 perpendicular to the horizontal axis of the figure or slanting them in the opposite direction. If two columns or more were distorted, a score of present was given.

A tally was made of the number of children for whom these distortions were scored present. Table 2 summarizes this data for each of the six groups.

ANALYSIS OF THE DATA

A relatively new method of data analysis, variously called multivariate information transmission analysis (McGill, 1954) and uncertainty analysis (Garner & McGill, 1956), was employed in the analysis of the data. This method makes possible the analysis of nominally scaled dependent variables in factorial design formats and can be used as a nonmetric and nonparametric analogue of the analysis of variance (Attneave, 1959; Garner & McGill, 1956). For designs in which the effects of orthogonal independent variables on a dependent variable are to be studied, the significance of interactions as well as main effects can be determined. The functions studied are amounts of "information transmitted" (or degree of uncertainty reduction) measured in "bits," for either main effects (T') or interactions (A'). The product of a T' or A' value with the constant 1.3863 N is distributed as chi square for the appropriate number of degrees of freedom (df). A bias in the resulting chi squares may be corrected by subtracting from it the df for that source (Attneave, 1959),³ as can be seen in Table 3.

The amount of information transmitted from each source, its df , and the resultant

³ Attneave (1959) suggests the correction of the T' or A' value for bias. It is, however, readily demonstrable that the method used here is equivalent, and for the purpose of significance testing, computationally simpler.

TABLE 2
INCIDENCE OF BENDER GESTALT TEST DISTORTIONS FOR EACH DIAGNOSTIC GROUP

Diagnostic Groups	Distortions									
	Angulation		Rotation		Primitivation		Separation		Slant	
	Present	Absent	Present	Absent	Present	Absent	Present	Absent	Present	Absent
RY	18	2	4	16	7	13	9	11	13	7
RO	15	5	2	18	4	16	5	15	6	14
EY	15	5	2	18	4	16	5	15	12	8
EO	12	8	2	18	2	18	4	16	5	15
NY	11	9	3	17	5	15	4	16	6	14
NO	11	9	4	16	5	15	2	18	7	13

TABLE 3

MULTIVARIATE INFORMATION TRANSMISSION ANALYSIS OF THE EFFECTS OF AGE, DIAGNOSTIC GROUP, AND TYPE OF DISTORTION ON THE PRESENCE-ABSENCE OF DISTORTION

Source of Transmission	Amount (Bits)	df	Chi Square	<i>p</i>	Corrected Chi Square	<i>p</i>
T' (Age; P-A)	.00917	1	7.627	<.01	6.627	<.02
T' (Diagnostic Group; P-A)	.00887	2	7.378	<.05	5.378	.07
T' (Type Distortion; P-A)	.11844	4	98.516	<.01	94.516	<.01
A' (Age, Diagnostic Group, P-A)	.00497	2	4.134	<i>ns</i>	2.134	<i>ns</i>
A' (Age, Type Distortion, P-A)	.00380	4	3.161	<i>ns</i>	—	<i>ns</i>
A' (Diagnostic Group, Type Distortion, P-A)	.01126	8	9.366	<i>ns</i>	1.366	<i>ns</i>
A' (Age, Diagnostic Group, Type Distortion, P-A)	.00600	8	4.991	<i>ns</i>	—	<i>ns</i>
T' (Age, Diagnostic Group, Type Distortion; P-A)	.16251	29	135.173	<.01	106.173	<.01

chi square are additive. The over-all sum reflects the total amount of information transmitted between the independent variables taken together and the dependent variable or the amount of uncertainty reduction, or more generally still, the association. It is analogous to a multiple correlation with the dependent variable, when all variables are metric.

RESULTS

As applied to this problem, age (the three younger subgroups compared with the three older subgroups), diagnostic group (comparisons among Groups R, E, and N), and type of distortion are the independent variables and presence or absence of the distortion (P-A) is the dependent variable. Table 3 sets forth these results.

The highly significant total amount of information transmitted indicates that the joint effect of age, diagnostic group, and type of distortion is a real one. This effect may, however, be trivial since the effect of type of distortion is merely due to the fact that the five distortions investigated are of different incidence in the population. It is in fact this source that accounts for the lion's share of the total amount of information transmitted. None of the interactions (A') even approach significance. The data thus fails to indicate the existence of any patterns of type of distortion as a function of age, or of diagnostic group, or of age and diagnostic group.

The analysis of the data does indicate, however, that the age difference contributes significantly to the presence of the distortions, the younger groups providing the greater num-

ber of distortions. Since the two age groups studied differed by about two years, the significance of the difference between them attests to the ability of the Bender gestalt test to differentiate along a chronological age continuum. Thus the five distortions investigated seem to be affected by maturational factors between ages 8 and 12. This finding confirms Bender's descriptive data (1938) that the test is sensitive to differences in perceptual-motor ability up to the age of 12 years.

The differences among Groups R, E, and N meet the .05 criterion of significance for the uncorrected chi square. This would suggest group differences in the incidence of the distortions. However, upon correction the significance of the difference slips to the .07 level. It was deemed worthy to follow this up and to attempt to identify the source of the difference more specifically. Accordingly, further complete multivariate transmission analyses were performed on the diagnostic groups taken a pair at a time. In these analyses, it was found that total distortion was significantly greater in the retarded readers than in the normal group ($T' = .01239$, corrected chi square = 5.870, $p < .02$), and tended to be greater in the retarded readers than in the emotionally disturbed group ($T' = .00780$, corrected chi square = 3.325, $p < .07$). In these supplementary analyses, the interactions also failed to approach significance.

The presence of these distortions readily distinguishes children with reading disabilities from normal children. However, when emotionally disturbed, normal readers are compared with reading retarded children, the

distortions do not distinguish so efficiently. The difference here, just below significance, suggests that though reading retardation may well reflect immaturity in perceptual-motor development, this developmental hypothesis cannot be used to account in full for the retardation. Emotionally disturbed, normal readers offered these distortions too frequently to permit their exclusive association to reading retardation.

The findings of the present study lend support to the importance of the developmental concept in understanding reading disability. The question of the etiology of this developmental lag should still be raised. The present study can offer no evidence on this point.

SUMMARY AND CONCLUSIONS

The relationship between perceptual-motor development and reading disability was investigated. The Bender gestalt test was administered to children retarded in reading ability, children defined as emotionally disturbed but normal readers, and normal children. These groups were matched on a number of pertinent variables.

It was hypothesized that the following Bender gestalt test signs would discriminate among the three diagnostic groups and between two age levels within these groups: (a) difficulty in constructing angles, (b) rotation of figures, (c) primitivation of figures, (d) separation of adjacent parts of figures, (e) inability to maintain slant of figure.

The presence or absence of these distortions, the dependent variable in a four dimensional information transmission analysis, was affected by the type of distortion and by the age variable. The younger Ss produced more distortions than the older group.

A suggestive finding regarding the frequency with which the distortions occurred in the three diagnostic groups was explored. The distortions were indeed offered more frequently by the reading disability children than by the normal children. When the reading retarded and the emotionally disturbed children were compared, the significance of the difference fell just below an acceptable level.

The developmental hypothesis has received some support from these findings though it

cannot claim to explain fully the presence of reading disorders. No etiological implications can be drawn.

REFERENCES

- ATTNEAVE, F. *Application of information theory to psychology*. New York: Holt, 1959.
- BENDER, L. *A visual-motor gestalt test and its clinical use*. New York: Amer. Orthopsychiat. Ass., 1938.
- BENDER, L. Psychological principles of the visual-motor gestalt test. *Trans. NY Acad. Sci.*, 1949, 11, 164-170.
- DE HIRSCH, K. Gestalt psychology as applied to language disturbances. *J. nerv. ment. Dis.*, 1954, 120, 257-261.
- FABIAN, A. Vertical rotation in visual-motor performance: Its relation to reading reversals. *J. educ. Psychol.*, 1945, 36, 129-154.
- GARNER, W. R., & MCGILL, W. J. Relation between uncertainty, variance, and correlation analyses. *Psychometrika*, 1956, 21, 219-228.
- JARVIS, V. Clinical observations on the visual problem in reading disability. In R. Eissler et al. (Eds.), *The psychoanalytic study of the child*. Vol. 13. New York: Int. Univer. Press, 1958.
- KUHLMANN, F., & ANDERSON, R. *Kuhlmann-Anderson Intelligence Test master manual*. Princeton: Personnel Press, 1952.
- MCGILL, W. J. Multivariate information transmission. *Psychometrika*, 1954, 19, 97-116.
- PASCAL, G. R., & SUTTELL, B. J. *The Bender-gestalt test: Quantification and validity for adults*. New York: Grune & Stratton, 1951.
- PASTOVIC, J. J., & GUTHRIE, G. M. Some evidence on the validity of the WISC. *J. consult. Psychol.*, 1951, 15, 385-386.
- SILVER, A. A. Postural and righting responses in children. *J. Pediat.*, 1952, 41, 493-498.
- SILVER, A. A. Diagnosis of the various syndromes encountered in retarded pre-adolescent children. In *The pre-adolescent exceptional child*. Woods Schools: Proc. 35th Conf. Child Research Clinic, 1953.
- TERMAN, L. M., & MERRILL, M. A. *Measuring intelligence*. New York: Houghton Mifflin, 1937.
- TIEGS, E. W., & CLARK, W. W. *California Reading Test*. (Primary form.) Los Angeles: California Test Bureau, 1950. (a)
- TIEGS, E. W., & CLARK, W. W. *California Reading Test*. (Elementary form.) Los Angeles: California Test Bureau, 1950. (b)
- TRAXLER, A. E. Comparison between IQ's on the new edition of the Kuhlmann-Anderson Intelligence Tests and Binet IQ's. In, 1940 fall testing program in independent schools and supplementary studies. *Educ. rec. Bull.*, 1941, No. 31.
- VORHAUS, P. A. Non-reading as an expression of resistance. *Rorschach res. Exch.*, 1946, 10, 60-69.
- WECHSLER, D. *Manual, Wechsler Intelligence Scale for Children*. New York: Psychological Corp., 1949.

(Received October 2, 1959)

AN EXPLORATION OF THE RELATIONSHIP BETWEEN HYPNOTIZABILITY AND ANXIETY AND/OR NEUROTICISM¹

FRED HEILIZER²

University of Rochester

Studies by Eysenck (1943b, 1944, 1947) and Himmelweit, Desai, and Petrie (1946) indicate the existence of a positive correlation between neuroticism and hypnotizability. In the most comprehensive of these studies, Eysenck (1947) found a perfect positive rank order correlation between neuroticism and postural sway for groups of neurotics and normals. In the same publication, Eysenck (1947), on the basis of a 47-item questionnaire, demonstrated that "suggestible neurotics and nonsuggestible neurotics differ with precisely the same factor [neuroticism] as do neurotics as a whole and normals" (p. 272). In another study, Eysenck (1944) reports that the neurotics swayed significantly more than the normals, and concludes "that *there is a close relation between primary suggestibility and neurosis*" (p. 410). In a third study, Eysenck (1943b) found that neurotics were not responsive on the postural sway test when they were first admitted to the hospital and that their postural sway scores decreased during a period of four weeks at the hospital. He interprets this finding as supporting his hypothesis that neuroticism and hypnotizability are positively related by assuming that hospitalization resulted in a psychological improvement in the patients. Himmelweit et al. (1946) found a tetrachoric correlation of .51, based upon a dichotomy between neurotic and normal subjects (Ss), between postural sway and psychiatric (neurotic) diagnosis.

In a second group of studies no relation-

ship between neuroticism and hypnotizability was found. Davis and Husband (1931), Bartlett (1936b), and Messer, Hinckley, and Mosier (1938) report no relationship between neuroticism, as defined by the Thurstone or Bernreuter Personality Schedule, and hypnotizability, as defined by a depth of hypnosis scale or by postural sway. Bartlett (1936a), using hospitalized neurotics and normals, found no difference between the two groups on the postural sway test. On the basis of data collected from hospitalized neurotic soldiers, Eysenck (1943a), in an early study, concludes that there is "good agreement with the findings of Bartlett, who found no differences between neurotics and normals in the Body Sway Test" (p. 30). Ingham (1954) reports that two of Eysenck's students found that neurotics and normals did not differ significantly on the postural sway test.

Despite the general superiority of the first group of studies, in terms of the *N* utilized and the specification of the neuroticism dimension, there are two sets of findings which indicate that the conclusion drawn by Eysenck and Himmelweit et al. (that there is a general positive relationship between neuroticism and hypnotizability) may be premature. Benton and Bandura (1953) report a nonsignificant correlation ($r = .19$) between static ataxia (a neuroticism indicator) and postural sway when utilizing normal Ss. This is consistent with a rarely quoted datum ($r = .06$ between static ataxia and postural sway with normal Ss) reported by Eysenck (1947, p. 277). The second set of findings is provided by Ingham (1954, 1955) who reports (a) that neurotics demonstrate significantly more postural sway and arm movement suggestibility than do normals, but (b) that

¹ This paper was abstracted from a dissertation submitted in partial fulfillment of the PhD degree. I would like to express my gratitude to G. R. Wendt and E. L. Cowen for their interest and assistance in this study.

² Now at the VA Hospital, Northampton, Mass.

there is no evidence of such a difference if baseline activity (without direct repetitive verbal suggestions) is controlled. Further, Ingham (1955) reports that neurotic Ss who had taken sedatives the night before scored significantly higher on the arm movement test than did neurotic Ss who had not taken sedatives. However, it is not clear on the basis of these data whether the hypnotizability scores are related to neuroticism *per se*, to the taking of sedatives *per se*, or to a combination of these factors.

PROBLEM

The purpose of this study was to investigate the relationship between neuroticism and hypnotizability among normals, utilizing (a) an adequate sample, (b) more precise measurement of the postural sway and static ataxia variables than is usually made, (c) the heat illusion test of hypnotizability in addition to postural sway, and (d) a sampling of measures of neuroticism and anxiety. The neuroticism and anxiety concepts were used interchangeably on the theoretical grounds that the two concepts are closely related (Fenichel, 1945; May, 1950; White, 1948). A supporting datum for such usage has been gathered by Franks (1954) who found a correlation of .92 between the Maudsley Medical Questionnaire, a widely used measure of neuroticism, and the Taylor anxiety scale. Since there is no exact counterpart of a psychiatric interview with normal college students and since the problem of selectively validating one or a few measures of anxiety or neuroticism has not yet been satisfactorily resolved, a variety of measures were used. Although such a procedure may not be conducive to an immediate decision to accept or reject the null hypothesis, it would seem to be more conducive to a fruitful line of investigation in these circumstances.

There are two methodological usages in the studies outlined above which are central to this study. (a) Hypnotizability has been almost invariably defined in terms of the postural sway test. In only one study (Davis & Husband, 1931) was hypnotizability defined in terms of a depth of hypnosis scale. (b) The studies which indicate the existence of a positive correlation between neuroticism and

hypnotizability defined neuroticism in terms of either static ataxia or the Maudsley Medical Questionnaire. In evaluating the relationship between neuroticism and hypnotizability among normal Ss, then, the most direct comparability involves postural sway as the independent variable and static ataxia and the Taylor anxiety scale (see above: Franks, 1954) as the dependent variables. The use of the heat illusion test of hypnotizability and the additional measures of anxiety and neuroticism constitute an attempt to extend the generality of this investigation.

INSTRUMENTS

Hypnotizability, the degree or depth to which a person can be hypnotized, was defined primarily in terms of reaction to the postural sway test (Furieux, 1952; Heilizer, 1959) and secondarily in terms of reaction to the heat illusion test (Eysenck & Furieux, 1945; Furieux, 1946; Furieux, 1952). The postural sway test was introduced to S as a test of motor activity. A control period and an experimental period of $2\frac{1}{2}$ min. each were utilized. S was instructed to relax and stand still during the control period. The experimental period consisted of direct repetitive suggestions to S that she was falling forward. Measurements of deviations from the baseline were made at 5-sec. intervals. The 30 measurements for each period were averaged and the difference between the two averages constituted the postural sway score. The heat illusion test was introduced as a measure of sensory threshold. S was shown how a small heating element became hot as a pointer was moved along a scale from 0 to 100. The heating element was then placed against S's forehead and she moved the pointer along the scale until she reported feeling the heat. A second trial was run with the heating element disconnected. If S reported feeling the heat on this trial, she was given a positive (hypnotizable) score.

The dependent measures were derived from five instruments or tasks: the control period of the postural sway task, the Biographical Inventory, the Bills-Vance-McLean Index of Adjustment and Values (the Bills), a paper-and-pencil mirror tracing task, and stories written in response to TAT cards.

The static ataxia measure (Eysenck, 1947; Himmelweit, et al., 1946) was derived as the standard deviation of the 30 measurements taken during the control period of the postural sway task. The A Scale (Taylor, 1953) and the Lie Scale (Welsh & Dahlstrom, 1956) were utilized from the Biographical Inventory. The self-concept-ideal-self discrepancy (Column IV) was used from the Bills (Bills, Vance, & McLean, 1951). Time (in seconds) and error scores were derived from a mirror tracing task which utilized a six-pointed star (Waters & Sheppard, 1952). The error scores were recorded in terms of 0.1" unit excursion amplitude by summing the largest single inside and outside excursion for each of the 12 straight-line segments to produce a total error score. The Discomfort Relief Quotient (DRQ) (Dollard & Mowrer, 1947) was scored in phrase units from stories written in response to three TAT cards (7BM, 12M, 6BM). In addition to the DRQ, an Emotionality scale, a Distance scale, three Constriction scales, and a Time/Phrase scale were utilized. The Emotionality scale was derived as the ratio of the number of phrases expressing emotion (relief or discomfort) to the total number of phrases. The Distance scale was based on the assumption that a person who typically reacts to anxiety-producing situations by withdrawing will, when writing a story in response to a TAT card, refrain from projecting himself into the story; he will write a story qualified by statements which detract from the projective quality. Such statements were classified under four headings:

1. Personal references (e.g., "I would say . . .;" "I think . . .;" "I believe . . .;" "As you can see . . .;")
2. Probability statements in which the outcome is undecided or doubtful (e.g., "Either . . . or . . .;" "Probably . . .;" "Perhaps . . .;" "Seems . . .;" "Appears to be . . .;" "Assuming . . .;")
3. Concrete statements (e.g., "The characters are . . .;" "The outcome is . . .;" "This is a picture of . . .;" "In the first slide . . .;" "In this he says . . .;")
4. Vague statements, referring to a vague "something" or to indefinite ages (e.g., "Something is bothering him . . .;" "He is think-

ing of something in his past . . .;" "He is 25 or 26 years old . . .;" "She is between 20 and 25 . . .;").

The score was derived as the ratio of the number of statements reflecting distance to the total number of phrases. Constriction scales were derived from the DRQ, Emotionality, and Distance scales by dividing the scores on the respective scales by the total number of phrases, thus giving added weight to the amount produced by reducing the scores according to the total number of phrases. This procedure was based on the assumption that one of the usual reactions of anxious Ss, when placed in a fairly unstructured situation, is to withdraw by producing as little as possible, that is, constriction. The Time/Phrase scale, utilized only for the second TAT testing (see below), was derived as a rate-of-production indicator: the ratio of the amount of time, in seconds, to the total number of phrases. Experimenter's (*E*'s) scoring reliability, based upon the same phrase units for each scoring of 50 randomly selected stories, was .95, .85, and .98 for the DRQ, Emotionality, and Distance scales, respectively.

SUBJECTS AND PROCEDURES

The Biographical Inventory was distributed to a class of 135 female introductory psychology students by the instructor with a request to complete it at home for purposes of standardization. Of the 135 students, 128 completed the Biographical Inventory, 75 volunteered for this experiment, and 62 completed the experimental situation. (Fifty nine of the 62 Ss completing the study also completed the Biographical Inventory.) Comparisons were made between the A Scale and Lie Scale scores of volunteers and nonvolunteers. The two groups were not significantly different (A Scale: $t = 0.26$; $F = 1.38$; Lie Scale: $t = 0.24$, $F = 1.06$). Thus, it appears that the Ss utilized in this study were a representative sample of the class of 135 students.

The Bills and the three TAT cards were presented in group sessions. A 6-min. time limit was allowed for each TAT card. The following tasks were then presented in individual sessions: postural sway, heat illusion, the three TAT cards, and mirror tracing. The

TAT cards were presented with no time limit to 49 randomly selected Ss. Originally, we had expected Card 12M to elicit many stories concerning hypnosis which could be analyzed as to content and outcome. However, only 15 stories about hypnosis were elicited during the group testing and many of these were only incidentally concerned with hypnosis. Therefore we repeated the TAT presentations with the incidental remark preceding presentation of Card 12M: "This is an interesting one. Did you know that almost everyone makes up a story about hypnosis for this card?" This resulted in 11 hypnosis stories, several of which were only incidentally concerned with hypnosis.

RESULTS AND DISCUSSION

The relationship between postural sway and each of the dependent variables was evaluated by Pearson and curvilinear correlations. (The curvilinear correlations were computed with postural sway as the independent variable.) Fisher z scores and phi coefficients were computed utilizing a 65th percentile division of the postural sway distribution (Heilizer, 1959). Six of the 76 tests of significance reached the .05 level: a Pearson correlation of .29 for Emotionality (second TAT testing); curvilinear correlations of .77 and .53 for mirror tracing time and error scores, respectively (in which the deviations from a zero, linear slope were confined to the high postural sway group—above the 65th percentile—and were in the shape of an inverted U); and phi coefficients for DRQ/Constriction (−.38), Emotionality/Constriction (−.30), and Time/Phrase (−.30), all of the second TAT testing.

Neither static ataxia nor the Taylor anxiety scale yielded a significant relationship with postural sway despite the use of a more precise measurement of the postural sway and static ataxia variables than had previously been made in the pertinent literature. We conclude, then, that the positive correlation between neuroticism and hypnotizability reported by Eysenck and Himmelweit et al. does not occur among normal Ss when hypnotizability and neuroticism are defined in terms of similar instruments.

If the remainder of the dependent vari-

ables are divided into measures with currency in the literature (Lie Scale, Bills, mirror tracing, DRQ) and measures which were derived for this study on rational grounds (Emotionality, Distance, the three Constriction scales, Time/Phrase) there are no distinctive trends in either group, with 2 out of 24 evaluations reaching significance in the former group and 4 out of 44 evaluations reaching significance in the latter group.

The relationship between heat illusion (27 positive and 35 negative responses) and each of the dependent variables was ascertained by the phi coefficient. Two of the 19 tests of significance reached the .05 level: the phi coefficients for Distance (.36) and Distance/Constriction (.29), both of the first TAT testing.

Thus, on the basis of the few significant relationships and the inconsistencies among the significant data, we conclude that the null hypothesis cannot be rejected for either the primary or secondary measures of hypnotizability or neuroticism and/or anxiety. It is possible that one or more of the significant results may reflect a true relationship. However, the assertion that one or more of the significant results does reflect a true relationship would be acceptable only upon replication.

It seems appropriate to conclude that the positive relationship between neuroticism and hypnotizability reported by Eysenck and Himmelweit et al. does not apply to normal Ss. At this point, Ingham's data, indicating that hypnotizability may not be related to neuroticism per se, assume added importance. It is possible that the relationship between neuroticism and hypnotizability among neurotics is the result of a drug usage which does not occur to an appreciable extent among normals. If this proves to be the case, then no simple relationship between neuroticism and hypnotizability can be stated, and the drug action might profitably become the focus of research interest.

SUMMARY

This study was designed to examine the relationship between hypnotizability and neuroticism among normal Ss, utilizing (a) an adequate sample, (b) more precise measure-

ment of the postural sway and static ataxia variables than is usually made, (c) the heat illusion test of hypnotizability in addition to postural sway, and (d) a sampling of measures of neuroticism and anxiety. Sixty-two college female volunteers completed the postural sway and heat illusion tests for hypnotizability and a variety of tests of neuroticism and anxiety.

The null hypothesis could not be rejected, thus leading to the conclusion that the positive relationship between neuroticism and hypnotizability reported to occur with neurotic Ss does not occur with normal Ss. It was speculated that the relationship between neuroticism and hypnotizability among neurotics is the result of a drug usage which does not occur to an appreciable extent among normals.

REFERENCES

- BARTLETT, M. R. Suggestibility in psychopathic individuals. *J. gen. Psychol.*, 1936, 14, 241-247. (a)
- BARTLETT, M. R. Relation of suggestibility to other personality traits. *J. gen. Psychol.*, 1936, 15, 191-196. (b)
- BENTON, A. L., & BANDURA, A. "Primary" and "secondary" suggestibility. *J. abnorm. soc. Psychol.*, 1953, 48, 336-340.
- BILLS, R. E., VANCE, E. L., & MCLEAN, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, 15, 257-261.
- DAVIS, L. W., & HUSBAND, R. W. A study of hypnotic susceptibility in relation to personality traits. *J. abnorm. soc. Psychol.*, 1931, 26, 175-182.
- DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, 42, 3-32.
- EYSENCK, H. J. Suggestibility and hysteria. *J. Neurol. Psychiat.*, 1943, 6, 22-31. (a)
- EYSENCK, H. J. Suggestibility and hypnosis: An experimental analysis. *Proc. Roy. Soc. Med.*, 1943, 36, 349-354. (b)
- EYSENCK, H. J. States of high suggestibility and the neuroses. *Amer. J. Psychol.*, 1944, 57, 406-411.
- EYSENCK, H. J. *Dimensions of personality*. London: Routledge & Kegan Paul, 1947.
- EYSENCK, H. J. *The scientific study of personality*. London: Routledge & Kegan Paul, 1952.
- EYSENCK, H. J., & FURNEAUX, W. D. Primary and secondary suggestibility: An experimental and statistical study. *J. exp. Psychol.*, 1945, 35, 485-503.
- FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
- FRANKS, C. M. An experimental study of conditioning as related to mental abnormality. Unpublished doctoral dissertation, Univer. London, 1954. (Reprinted: Eysenck, H. J. Reminiscence, drive, and personality theory. *J. abnorm. soc. Psychol.*, 1956, 53, 328-333.)
- FURNEAUX, W. D. The prediction of susceptibility to hypnosis. *J. Pers.*, 1946, 14, 281-294.
- FURNEAUX, W. D. Hypnotic susceptibility as a function of waking suggestibility. In L. M. LeCron (Ed.), *Experimental hypnosis*. New York: Macmillan, 1952. Pp. 115-136.
- HEILIZER, F. A note on the postural sway distribution. *J. consult. Psychol.*, 1959, 23, 493-495.
- HIMMELWEIT, H. T., DESAI, M., & PETRIE, A. An experimental investigation of neuroticism. *J. Pers.*, 1946, 15, 173-196.
- INGHAM, J. G. Body-sway suggestibility and neurosis. *J. ment. Sci.*, 1954, 100, 432-441.
- INGHAM, J. G. Psychoneurosis and suggestibility. *J. abnorm. soc. Psychol.*, 1955, 51, 600-603.
- MAY, R. *The meaning of anxiety*. New York: Ronald, 1950.
- MESSER, A. L., HINCKLEY, E. D., & MOSIER, C. I. Suggestibility and neurotic symptoms in normal subjects. *J. gen. Psychol.*, 1938, 19, 391-399.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- WATERS, R. H., & SHEPPARD, R. The mirror drawing experiment: A brief historical note. *J. gen. Psychol.*, 1952, 46, 63-72.
- WELSH, G. S., & DAHLSTROM, W. G. (Eds.) *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.
- WHITE, R. W. *The abnormal personality*. New York: Ronald, 1948.

(Received October 2, 1959)

PERSONALITY CHARACTERISTICS ASSOCIATED WITH RESISTANCE TO CHANGE¹

CARROLL E. IZARD

Vanderbilt University

A number of studies have shown that people misperceive, change their opinions, or otherwise behave so as to indicate that they are yielding or conforming in response to the pressure of the group, to authority, or to some other form of influence (Asch, 1952; Hovland, Janis, & Kelley, 1953; Hovland & Weiss, 1951; Sherif & Harvey, 1952). These researches have demonstrated that a number of social conditions will bring about such change. On the average, experimental subjects (Ss) exposed to one of these conditions will yield, misperceive, or change their stand on an issue significantly more than a control group. Since the individual members of a group—even an experimentally manipulated group—differ in important ways, it is not surprising that psychologists are beginning to study the personality differences among people who yield in varying degrees to outside influence (Boomer, 1959; McDavid, 1959). These studies have not yielded data on specific or well delineated personality variables. One reason for this was the lack of personality measures adequate to the task.

Recent research has pointed to Edwards' Personal Preference Schedule (EPPS) as a promising possibility. Bernardin and Jessor (1957), utilizing the EPPS to measure personality characteristics, tested the hypothesis that a group high on Autonomy and low on Deference would be less dependent than a group low on Autonomy and high on Deference. The Ss in each group were exposed to

experimental situations designed to elicit behavior relevant to three aspects of dependency: (a) reliance on others for approval, (b) reliance on others for help, and (c) conformity as measured in an Asch-type group situation. The high Autonomy, low Deference group was significantly less dependent in terms of the first two criteria but not on conformity. They thought that the Autonomy-Deference scales failed to predict conformity scores because of the difference between the behavioral situations involved in responding to the EPPS and the concrete reality of the Asch-type group situation. Gisvold (1958) attempted to check on this latter proposition. He modified Asch's method so that the pressure to conform was artificially produced—an S did not see or hear the actual responses of the rest of the group but was under the illusion that the three incorrect responses exhibited by the experimenter (E) were the opinions of the other three Ss in the experimental situation. Under these conditions Autonomy correlated significantly with conformity but Deference did not. Correlations for other EPPS scales and conformity were not reported in either of the foregoing studies.

In the present study an hypothesis relating to a third type of social influence was formulated for Autonomy and Deference as well as for Dominance and Abasement. Since the criterion measure was in terms of resistance to change in response to an authority figure's attempt to influence, it was hypothesized that Autonomy and Dominance would correlate positively with the criterion while Deference and Abasement would correlate negatively. The selection of these EPPS scales as predictors of behavior change in this experimental situation was largely on the basis of

¹ This study was financed in part by a Vanderbilt-ONR contract Nonr 2149(02), O. J. Harvey, Principal Investigator. The opinions and conclusions do not necessarily reflect those of the Navy Department. The author is indebted to Harvey for valuable suggestions, particularly with regard to the procedure for measuring the effects of influence.

the judged relevance of the item content of these scales to the predicted behavior. According to the item content of the scales: Autonomy relates to a person's preference for behaving in terms of his own thoughts and feelings independently of the thoughts and feelings of others; Deference relates to a person's preference for seeking the opinion and advice of others and for looking to others for decisions and leadership; Dominance relates to a person's preference for a superordinate role, for arguing for his own point of view; Abasement relates to a person's preference for giving in rather than fighting to have his own way and to a tendency to feel inferior to other people. The reports of Bernardin and Jessor and Gisvold, published subsequent to the completion of the present experiment, lend some support to the hypothesis. However, the former authors did not report the separate correlations of Autonomy and Deference with their criteria of dependency or conformity, and as already indicated the latter found no significant relation between Deference and conformity. Results favoring the present hypothesis will throw further light on the validity of the selected EPPS scales for predicting *Ss'* responses to interpersonal influence and will help identify the personality characteristics associated with tendency to change.

PROCEDURE

The 39 *Ss* were all the members of a class in psychology. They were given the EPPS during a class hour early in the semester. Following this, each *S* was asked to meet with *E* individually to participate in an experiment in judgment. On arrival at the laboratory, *S* was told that he would be taken into a dark room where he would estimate the distance between two points of light. *S* was seated in the dark room and given pencil and pad with which he was to record his estimates in inches. The stimulus lights were exhibited for .3 sec. with an interval of .4 sec. between them and with an interval of 5 sec. between each pair. They were approximately eye level, 15 ft. ahead of *S*. The horizontal distance between the points of light was fixed at 24 in. throughout the experiment. Each *S* made 120 judgments in all, with approximately 1 min. break between each set of 30. At the end of the first set, *S* was told:

You have just seen the standard series of lights which varied from — (S's minimum estimate) to — (S's maximum estimate) inches apart. I shall continue to show you lights within this range,

except that occasionally I'll show a pair of lights definitely farther apart than any in the standard series. For the rest of the experiment, you should say "standard" if the pair of lights is within the range — to — which you just saw. You should say "longer" if the pair of flashes are distinctly farther apart than any in the standard series.

S was then shown 90 more pairs of lights, each at the fixed 24 in. apart.

It was considered that in the process of making the first 30 estimates *S* formed a concept of the distance or range of distances between pairs of lights in the series. Thus, for each subsequent set of 30 judgments, the score for resistance to change of concept was taken as the largest number of consecutive "standard" responses. The final index of resistance to change of concept for a given *S* was obtained by summing the three resistance scores from the three sets of 30 judgments.

RESULTS

Table 1 shows the correlations of the four EPPS scales with the criterion of resistance to change. The correlations for men and women are reported separately, since the sexes have different means on the EPPS variables under consideration. Since a directional hypothesis was made, the one-tailed test was applied.

For the men, the correlations were all in the predicted direction. The correlations of Autonomy and Dominance with the criterion were significant at the .05 level of confidence and that for Abasement closely approached this level. The correlation of Deference and the criterion was in the predicted direction but failed to reach significance. For the women, none of the four correlations were significant at the .05 level. The correlation

TABLE 1
CORRELATIONS BETWEEN THE PERSONALITY (EPPS)
CHARACTERISTICS AND THE CRITERION OF
RESISTANCE TO CHANGE

	Resistance to Change	
	Males (<i>N</i> = 22)	Females (<i>N</i> = 17)
Autonomy	.38*	.05
Dominance	.38	-.13
Deference	-.24	-.34
Abasement	-.35	-.04

* The critical value for *r* at *p* = .05 by one-tailed test is .36 for men, .42 for women.

for Deference and the criterion was in the right direction and approached the magnitude of the correlations for the men.

For both sexes, correlations were computed for the 11 EPPS variables for which no hypotheses were made. A two-tailed p of .01 was required for these correlations, in view of the number computed and the lack of directional hypotheses. On this basis, only the correlation for men of $-.53$ between Order and the criterion approached significance. Only two other nonsignificant correlations were of the order of magnitude reached by the significant hypothesized correlations—for Succorance r was .49 for women and for Endurance r was $-.35$ for men.

DISCUSSION

It is difficult to compare these results with those of the two previous studies relating personality (EPPS) variables and change. Neither of them report separate correlations for men and women. Both studies utilized only the Autonomy and Deference scales and one (Bernardin & Jessor, 1957) used the two scales as a multiple criterion for grouping without giving any indication of the contribution of the separate scales to the prediction of hypothesized behavior. All three studies used different methods of exerting social influence or pressure to change.

If both Autonomy and Deference contributed to the differentiation of dependent and independent Ss in the Bernardin and Jessor study and if susceptibility to change is indeed a part of the dependency construct, then Autonomy is the one EPPS variable that had a measure of validity in all three studies. Bernardin and Jessor did not get a difference between their dependent and independent groups in an Asch-type situation, but Gisvold did when he modified the Asch technique so that Ss could conform without publicly announcing a response contrary to a concrete reality situation. In the present study Ss could likewise conform or be influenced in privacy.

The difference between the relationships of EPPS variables and the criterion for men and women could mean that the personality variables lead to different responses in the two sexes, that the experimental situation had a different effect, or that the interaction of per-

sonality and situational variables was different for males and females. It is entirely possible that judging the distance between lights in a dark room could have quite different meaning and different degrees of relevance for the sexes.

Although none of the correlations for the EPPS variables judged irrelevant to the criterion actually reached significance at the required level, two of them were between the .05 and .01 level and were equal to or larger than the hypothesized correlations. Further, the correlations for these variables were consistent in direction for both sexes. These variables and the correlations with the criterion for males and females, respectively, were: Order, $-.53$, $-.28$; and Succorance, .19, .49. It is conceivable that Ss characterized by high Order may have felt considerably disoriented and insecure in the relatively unstructured, ambiguous experimental situation. If so, it is reasonable that they should be more susceptible to outside influence—instructions which gave additional structure and order to the situation. A possible, though highly speculative, explanation for the correlations of Succorance with the criterion is that Ss high on Succorance perceive attempts to change or influence them as a threat to their need for acceptance, affection, and encouragement.

The interpretation of the four variables involved in the hypothesis of this study is simpler, at least for male Ss. Autonomy and Dominance have a degree of validity for predicting resistance to interpersonal influence as measured in the experimental situation. It can be said that each of these scales measures, in part, a facet of behavior on the autonomy-heteronomy dimension. The two previous studies support this generalization with respect to Autonomy. One of these studies, and the present one to a lesser degree, found that Deference also measured an aspect of behavior on this dimension for both men and women.

The fact that the male Ss varied in the extent to which they changed their perceptual-cognitive response to a stimulus situation under influence of E 's instruction and the fact that this change tended to be associated with certain operationally defined personality characteristics have implications for various inter-

personal situations. Such situations include some approaches to psychotherapy and counseling, where the advice, information, suggestion, or interpretation of one person is intended to influence another. Males low on Autonomy and Dominance and high on Abasement might be expected to change to a greater extent. There was some evidence that high Deference would contribute to susceptibility to change in both males and females.

SUMMARY

It was hypothesized that the personality characteristic of Autonomy, Dominance, Deference, and Abasement, as measured by the EPPS, would correlate with change of concept under influence of *E*'s instructions. The 39 Ss, all the students in one undergraduate class in psychology, were first given the EPPS. Ss were then taken individually to the laboratory where they were given the opportunity to form a concept of the distance between two stimulus lights exposed in a dark room. Following this, *E* gave instructions calculated to influence *S* to change his concept of the distance, although the actual distance remained the same.

For male Ss the hypothesis was confirmed for two of the four personality characteristics at less than the .05 level of confidence and for another at approximately the .05 level. Autonomy and Dominance correlated positively with resistance to change and Abasement correlated negatively. The correlation of Deference and resistance to change was not significant but was in the predicted direction for both men and women. The implications of this finding for the validity of the EPPS and for certain interpersonal interactions, including psychotherapy and counseling, were discussed briefly.

The hypothesis was not confirmed for fe-

male Ss. It was considered a possibility that the differences in correlations for men and women were artifacts relating to the experimental situation.

None of the 22 correlations for the EPPS variables not included in the hypothesis was significant at the required .01 level. However, two of the variables had correlations which were between the .05 and .01 levels and which were corroborated by correlations in the same direction for the other sex. Tentative interpretations of the meaning of these variables in relation to the experimentally evoked behavior were offered.

REFERENCES

- ASCH, S. E. Effects of group pressures upon the modification and distortion of judgments. In G. E. Swanson, T. N. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1952.
- BERNARDIN, A. C., & JESSOR, R. A construct validation of the Edwards Personal Preference Schedule with respect to dependency. *J. consult. Psychol.*, 1957, 21, 63.
- BOOMER, D. S. Subjective certainty and resistance to change. *J. abnorm. soc. Psychol.*, 1959, 58, 323-328.
- GISVOLD, D. A validity study of the Autonomy and Deference subscales of the EPPS. *J. consult. Psychol.*, 1958, 22, 445-447.
- HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion: Psychological studies of opinion change*. New Haven: Yale Univer. Press, 1953.
- HOVLAND, C. I., & WEISS, W. The influence of source credibility in communication effectiveness. *Publ. opin. Quart.*, 1951, 15, 633-650.
- MCDavid, JOHN, JR. Personality and situational determinants of conformity. *J. abnorm. soc. Psychol.*, 1959, 58, 241-246.
- SHERIF, M., & HARVEY, O. J. A study of ego functioning: Elimination of stable anchorages in individual and group situations. *Sociometry*, 1952, 15, 272-305.

(Received October 8, 1959)

CONSTRUCT VALIDITY OF THREE MASCULINITY-FEMININITY TESTS

GORDON A. BARROWS

Indiana Division of Mental Health, Indianapolis

AND

MARVIN ZUCKERMAN¹

Institute of Psychiatric Research, Indianapolis

Masculinity-Femininity (M-F) tests are usually developed using items which differentiate men from women. Obviously, the construct validity of these tests cannot rest entirely on this differentiation since much simpler criteria are available for this purpose. The construct validity of these tests must also be based on relationships with other variables within a sex group. Since there are several widely used tests of M-F, they should be highly intercorrelated if they are measuring the same construct. These tests should correlate with interest patterns which the society defines as masculine or feminine. There is no a priori reason to expect these tests to correlate with general intelligence; however, since men are usually found to be somewhat superior on tests involving quantitative or arithmetic ability, masculinity in men might correlate with these tests.

On the basis of theory and clinical lore, one might expect to find certain differences related to marital status variables. On the assumption that a strong masculine motive impels one toward the masculine role in marriage and parenthood; (a) married men might be expected to have higher masculinity scores than single men, (b) men who marry at a younger age might be expected to have higher masculinity scores than men who marry at an older age, (c) men who have more children might be expected to have higher masculinity scores than men who have fewer children.

Psychoanalytic theory suggests that men who marry women older than themselves are looking for a "mother-substitute." Such men might be expected to be lower in masculinity

than men who marry women younger than themselves, since the latter supposedly have resolved their "Oedipus Complex" through identification with their father.

Jung (1939) has suggested that the feminine component in men ("anima") becomes stronger as they age. This hypothesis is supported by data of Terman and Miles (1936) who found a decline in masculinity scores in males after high school. One might expect somewhat similar results on the M-F tests used in this study.

Most of the aforementioned hypotheses are highly speculative and negative evidence would not be crucial for either the theories or the validity of the tests. Positive evidence, however, would support the theories and test validity, particularly if more than one M-F test yields results in the expected direction.

SUBJECTS

The subjects (Ss) in this study were 2,296 males employed at "white-collar" jobs in a large Canadian business institution who were administered a battery of tests in connection with a program of personnel assessment. This group was above the general Canadian population in educational achievement. One percent completed elementary school; 38% had one to three years of high school; 39% had four years of high school; 19% had five years of high school; and 3% had one to four years of college. The age of the group ranged from 17 to 56 (see Table 4 for the distribution).

TESTS

The tests administered to the group included intelligence, aptitude, achievement, and

¹ Now at Department of Psychology, Brooklyn College.

personality tests. The analyses in this study will concern only the following tests:

1. *Masculinity-Femininity tests.* (a) Guilford-Zimmerman Temperament Survey (G-Z) M-F scale, (b) Minnesota Multiphasic Personality Inventory (MMPI) M-F scale, (c) the Strong Vocational M-F scale.

2. *Intellectual Abilities.* (a) Otis Quick-Scoring Mental Ability Test, (b) Watson-Glaser Critical Thinking Appraisal, (c) Thurstone Mental Alertness Test, (d) Cooperative English Reading Comprehension Test, (e) Cardall Arithmetical Reasoning Test.

3. *Vocational Interest.* The Kuder Preference Record—Vocational Form C.

MARITAL STATUS VARIABLES

The following marital status variables were analyzed in relation to M-F tests: (a) single vs. married status (excluding widowed and divorced Ss); (b) number of children; (c) wife's age in relation to husband's (older, younger, or same); (d) age of S at marriage.

One further variable analyzed in relation to M-F was the S's age.

RESULTS

Correlations between Measures of M-F

The correlations between the measures of M-F are contained in Table 1. In this table and all others giving correlations between the MMPI M-F test, the correlations involving the MMPI M-F are reversed in sign because this test is scored in the feminine direction while the G-Z and Strong M-F tests are scored in the masculine direction. All three

TABLE 1

INTERCORRELATIONS AMONG GUILFORD-ZIMMERMAN, MMPI, AND STRONG MASCULINITY-FEMININITY SCALES*

(N = 2296)

	MMPI	Strong
G-Z	.31	.34
MMPI		.33

* All three correlations are significant below the .001 level.

correlations are significant and of about the same magnitude (.3).

Correlations between the G-Z and Strong M-F scales and the other scales contained in these tests were available. These correlations are of some interest since they enable one to evaluate the validity of the correlation between M-F scales using the multitrait-multimethod system suggested by Campbell and Fiske (1959). In this case, however, there is only one trait being evaluated, M-F. These authors list several requirements for validity:

1. The validity correlation between two variables using different techniques to measure the same trait should be significantly different from zero and large enough to encourage further examination of validity. The correlation of .34 between the two M-F scales was highly significant but its magnitude is questionable. Considering the large N involved in the correlation there is some justification in proceeding to the next criterion.

2. The validity correlation should be higher than the correlations between the variables and other variables that have neither construct nor method in common. The .34 correlation between the M-F scales is higher than the correlations of the Strong M-F with 9 other G-Z scales and higher than the correlations of the G-Z M-F with 10 other Strong scales. The data satisfied the second criterion.

3. The validity correlation should be higher than correlations between the variables and other variables which are designed to get at different traits but happen to use the same method. The .34 correlation between M-F scales is lower than one of the nine correlations between G-Z M-F and other G-Z scales. G-Z M-F correlated .35 with the Objectivity scale. The validity correlation is lower than 2 of the 10 correlations between Strong M-F and the other Strong scales. Strong M-F correlated .55 with the Production Manager scale and .43 with the Certified Public Accountant, Senior scale. The data do not satisfy the last criterion although the number of correlations indicating the greater influence of method than construct is not excessive. Furthermore, it should be noted that the scales on the Strong were developed from the same pool of items so that correlations between scales may be a function of item overlap.

TABLE 2

CORRELATION BETWEEN MEASURES OF MENTAL ABILITY AND MEASURES OF MASCULINITY

(N = 2296)

Ability Tests	G-Z	Strong	MMPI
Otis S-A	.21*	.08*	-.14*
W-G Critical Thinking	.20*	.05	-.15*
Thurstone (Ling.)	.22*	-.02	-.22*
Thurstone (Quant.)	.17*	.12*	-.04
Thurstone (Total)	.23*	.04	-.18*
Coop. Reading Comp.	.24*	-.03	-.26*
Cardall Arith. Reasoning	.21*	.11*	-.07*

* Significant at or below .001 level.

Correlations of M-F with Measures of Ability

The G-Z M-F correlates positively and significantly with all measures of ability, verbal and quantitative (see Table 2). The Strong M-F correlates positively and significantly (using the conservative .001 level) with the Otis, the Thurstone Quantitative subtest, and the Cardall Arithmetical Reasoning Test. Unlike the previous two tests, masculinity on the MMPI correlates negatively with all ability measures and all the correlations are significant except the correlation with the Thurstone Quantitative subtest.

Correlations of M-F with Vocational Interests

Correlations of the three measures of M-F with the scales of the Kuder interest scales can be seen in Table 3. The general pattern of these correlations is nearly the same on all

TABLE 3

CORRELATION BETWEEN MEASURES OF INTEREST AND MEASURES OF MASCULINITY

(N = 2296)

Kuder Interest Scales	G-Z	Strong	MMPI
Outdoor	.14*	.18*	-.07*
Mechanical	.22*	.56*	.26*
Computational	.08*	.14*	.13*
Scientific	.25*	.42*	.21*
Persuasive	-.02	-.09*	.06
Artistic	-.12*	-.15*	-.12*
Literary	.01	-.29*	-.27*
Musical	-.23*	-.29*	-.23*
Social Service	-.04	-.15*	-.03
Clerical	-.15*	-.07*	.03

* Significant at or below .001 level.

TABLE 4

AGE AND MASCULINITY-FEMININITY

Ages	N	G-Z (Masculinity)	Strong (Masculinity)	MMPI (Femininity)
17-26	1315	4.84	45.3	23.76
27-36	383	5.07	45.1	24.45
37-46	468	4.87	43.7	25.22
47-56	128	4.82	43.8	24.38
	<i>F</i>	1.60	3.99	11.17
	<i>p</i>	<i>ns</i>	.01	.001

three measures. Mechanical, Computational, and Scientific interests correlate positively with masculinity; while Artistic, Literary, Musical, and Clerical interests tend to correlate negatively with masculinity. Outdoor interests correlate positively with masculinity on the G-Z and Strong scales but negatively with masculinity on the MMPI.

Age and M-F

Ss were grouped in four age groups as listed in Table 4. The M-F scores in these groups were compared using single-classification analyses of variance. Significant *F* ratios between age groups were found on the Strong and MMPI M-F scales. Looking at the group means on the Strong, it is apparent that the two younger groups (ages 17-36) are higher in masculinity than the two older groups (ages 37-56). The relationship is less clear on the MMPI where femininity scores increase from the youngest to the 37-46 age group and then decrease in the 47-56 age group.

TABLE 5

DIFFERENCES BETWEEN SINGLE AND MARRIED MEN ON MASCULINITY-FEMININITY SCALES

	G-Z (Masculinity)	Strong (Masculinity)	MMPI (Femininity)
Single (N = 1039)			
<i>M</i>	4.68	44.70	23.86
<i>SD</i>	1.90	10.00	5.46
Married (N = 1242)			
<i>M</i>	5.06	44.90	24.45
<i>SD</i>	1.81	8.70	4.40
<i>CR</i>	4.90	.50	2.81
<i>p</i>	.001	<i>ns</i>	.005

TABLE 6
AGE AT MARRIAGE AND MEASURES
OF MASCULINITY-FEMININITY

Age at Marriage	N	G-Z (Masculinity)	Strong (Masculinity)	MMPI (Femininity)
Under 21	60	5.15	46.5	23.60
21-25	452	5.22	45.7	24.07
26-30	416	5.00	44.6	24.60
31 +	313	4.91	44.1	24.99
	<i>F</i>	2.12	2.97	3.60
	<i>p</i>	<i>ns</i>	.05	.05

Marital Status and M-F

The married group was significantly higher than the single group on masculinity as measured by the G-Z, but they were higher on femininity as measured by the MMPI (see Table 5).

Number of Children and M-F

Married Ss were divided into five groups: no children, one child, two children, three children, and four or more children. Single classification analyses of variance did not yield significant *F* ratios on any of the three M-F scales.

Wife's Age in Relation to Husband's and M-F

A comparison of the three groups, wife older, wife younger, and wife same age, did not yield significant *F* ratios on any of the three M-F scales.

Age-at-marriage and M-F

The Ss were divided into four groups on the basis of their age at time of marriage (see Table 6). The *F* ratios between groups were significant (.05 level) for the Strong and MMPI M-F scales. On both scales, the relationship is similar: going from the younger age at marriage to the older age at marriage, one sees a decrease in mean masculinity scores (Strong) and an increase in mean femininity scores (MMPI).

DISCUSSION

The correlations between the three tests which purport to measure the trait masculinity-femininity are low, albeit highly significant. Nance (1949), who intercorrelated the same tests using a sample of male students in a college of education, found two of the correlations to be somewhat higher (G-Z and

MMPI, .43; G-Z and Strong, .28; Strong and MMPI, .51). Considering the size of our sample the reliability of the obtained correlations is high. A question does arise about possible national, vocational, and educational differences between populations. Considering only the results obtained on the population sampled in this study, it is apparent that although the three tests have some communality, the major part of their variance is not accounted for by the common factor. This is apparent in other aspects of the results, i.e., the finding that the G-Z and MMPI M-F tests give opposite results in correlations with ability measures and in comparisons of married and single groups.

Application of the Campbell and Fiske criteria to the Strong and G-Z tests indicated weak construct validity for M-F as measured by these tests. Considering the correlations with vocational interests it is apparent that all the tests relate in small but significant degree to masculine and feminine interest patterns. A glance at the M-F tests reveals that a portion of their items are stated in terms of vocational interests.

Masculinity on the G-Z is positively correlated with all measures of ability. The relationships are no more pronounced on quantitative than on verbal tests. Masculinity on the Strong seems to be more highly related to quantitative tests of ability in conformance with the stated hypothesis. Abilities on the MMPI are negatively correlated with masculinity. The results are less pronounced on the two quantitative tests. The results on the MMPI can be compared with those of Goodstein (1954) who found that male college students score high on femininity relative to the general population. A number of items on the MMPI M-F score cultural interests as feminine. An interest in poetry or the theater is feminine while an interest in sports or hunting is masculine. Similarly, a wish to be a librarian or journalist is feminine while a wish to be a soldier or contractor is masculine. While an increased interest in certain activities and vocations stemming from education or intelligence may indicate greater femininity (in the stereotype of the general population) it is doubtful whether this is a meaningful trait in the more educated. It is

not clear why masculinity on the G-Z is positively related to measures of ability other than quantitative ability. The G-Z often pairs vocations or interests of equal intellectual level in the items, e.g., "You would rather study mathematics and science than literature and music, T or F." Perhaps forcing a choice makes the more intelligent males subordinate general cultural interests to specific vocational interests which require high intelligence.

The only one of the marital status variables giving consistent results on at least two tests was age-at-marriage. Those who married when young had higher masculinity scores on the Strong and MMPI than those who married when older. As part of the larger study a "success index" was computed for every S in terms of average annual increments in salary over the period of employment with corrections for the change in the value of the dollar. The mean success index values for the four age-at-marriage groups were as follows: under 21, 24.3; 21-25, 19.7; 26-30, 13.0; over 31, 9.1. Differences between all groups were statistically significant and indicated that those men who married earlier in life had been more successful in this organization. In view of the association of age-at-marriage with both the success index and Strong and MMPI M-F, one might expect some relationship between masculinity and success. These correlations are low, .11 and .09, but highly significant ($p < .001$). Evidence from other tests related to the success index indicated that the more active, less restrained men had greater success. The aggressive, impulsive aspect of masculinity may be what leads to earlier marriage in highly masculine men. These men may find it easier to approach and establish relations with members of the opposite sex.

The evidence on the relation between age and masculinity is ambiguous. Although the Strong scale indicates a sharp drop in masculinity scores after 36 years of age, the MMPI indicates a drop in masculinity up to 46 and then an increase in the last age group (47-56).

It seems apparent to the investigators that masculinity-femininity is not a clearly defined construct. The tests used contain a mixture of interests and emotional attitudes. In-

terest factor analyses might help clarify the construct that the tests purport to measure.

SUMMARY

The Guilford-Zimmerman, Strong, and MMPI Masculinity-Femininity tests were given to 2,296 employees of a Canadian business institution along with tests of mental abilities and vocational interests. The relationships between these tests, and the relationships between masculinity and certain marital status and age variables were investigated. The correlations between the three M-F tests were low, but highly significant. Masculinity as measured by the G-Z and Strong tended to correlate positively with abilities, particularly quantitative ability. Masculinity on the MMPI correlated negatively with the ability tests, particularly the nonquantitative ones. All tests correlated significantly with interest scales. Mechanical, Scientific, and Computational interests correlated positively with masculinity while Artistic, Clerical, Musical, and Literary interests tended to correlate negatively with masculinity. Married Ss scored higher on masculinity on the G-Z, but single Ss were higher on the MMPI. Number of children and the relation of the wife's age to the husband's did not yield significant differences. Age-at-marriage yielded significant differences between four groups on the Strong and MMPI scales. Those married at younger ages tended to have higher masculinity scores. S's age yielded significant differences between groups on the Strong and MMPI M-F scales with a tendency for masculinity to drop with increasing age, but the trends were not completely linear.

REFERENCES

- CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, **56**, 81-105.
- GOODSTEIN, L. D. Regional differences in MMPI responses among male college students. *J. consult. Psychol.*, 1954, **18**, 437-441.
- JUNG, C. G. *The integration of the personality*. New York: Farrar & Rinehart, 1939.
- NANCE, R. D. Masculinity-femininity in prospective teachers. *J. educ. Res.*, 1949, **42**, 658-666.
- TERMAN, L. M., & MILES, C. C. *Sex and personality*. New York: McGraw-Hill, 1936.

(Received October 8, 1959)

AGGRESSION AND THE PICTURE-FRUSTRATION STUDY¹

J. KASWAN,

University of California, Los Angeles

M. WASMAN, AND LAWRENCE ZELIC FREEDMAN

Yale University

The Rosenzweig Picture-Frustration Study (P-F) is widely used as a clinical and research measure of verbal aggression, although its validity has not been clearly established. Most validating studies have attempted to correlate a specific criterion of aggression with the P-F scoring term "extrapunitive" (E)—defined as the percentage of responses in which the individual turns his aggression outward, against the environment. For example, Fry (1949), Holzberg and Hahn (1952), Vane (1954), and Weinberg (1952) have all found that juvenile delinquents or adult criminal offenders fail to manifest more E on the P-F than nonantisocial subjects (Ss), and have therefore concluded that the P-F is not valid. That such studies may represent too narrow an approach to validation is indicated by the APA Committee on Test Standards (1954, pp. 14-15) who note that test validity in the case of most clinical instruments must be evaluated by integrating evidence from many different sources and that often no single criterion measure or composite criterion can be identified. Thus aggression may occur in many different degrees, in many different situations, and in many different forms. Conversely, similar degrees of aggression could be expressed in different ways. It therefore may not be reasonable to expect that any particular measure of aggression which is arbitrarily selected as *the* single and specific criterion

out of the total spectrum of aggressive behavior should relate to P-F performance.

This paper represents an attempt to investigate the validity of the P-F by relating the quantity and intensity of aggressive responses obtained on it to aggression as described by various other techniques. We assume that if the P-F does refer to aggression it should relate to a substantial proportion of the other measures of aggression.

A measure of the *intensity* of aggressive reaction to frustration was designed in the belief that mild verbal aggression, such as "I don't agree," might characterize a different reactive process to frustration than that reflected by responses containing references to violent, uncontrolled aggression, such as "I'll break your neck." The failure of the present P-F scoring system to take into account such variations in intensity has been pointed out by Holzberg and Hahn (1952). This additional measure is expected to broaden the range of relationships between the P-F and other measures of aggression.

The investigation utilizes part of the data collected in an interdisciplinary study of antisocial offenders classified as sexual, aggressive, or acquisitive. Psychiatrists, psychologists, sociologists, and to a limited degree, lawyers, cooperated in obtaining data by the techniques appropriate to each discipline in such a way as to allow comparable statistics and other analyses. A great deal of information was thus compiled for each individual, providing possibilities for describing and relating behavior in a variety of contexts.

PROCEDURE

Subjects

The sample tested consisted of 121 male state prison inmates selected so as to match the total prison population on eleven variables, including age,

¹ A version of this paper was presented at the Eastern Psychological Association Meeting at Philadelphia, April 12, 1958.

This study was carried out as part of a research project on nonconformist behavior conducted within the Department of Psychiatry, Yale University, directed by Lawrence Z. Freedman, and supported by the State of Connecticut and Foundations Fund for Research in Psychiatry. The orientation of the paper does not necessarily reflect the theoretical framework of the over-all study.

education, and marital status. In terms of general characteristics the sample may be described as follows: 41 sexual offenders, 40 aggressive offenders, 40 acquisitive offenders; 72% white, 28% Negro; median age 36.4 years; 55% did not have any education beyond the 8th grade, 41% had some high school, 4% some college; median IQ was 96.2; median time in prison to date of testing was 2.2 years.

Ss were assured that participation or nonparticipation could have no possible effect on their prison status. Cooperation was voluntary. Approximately 20% of those called refused psychological testing. On the basis of the freedom of expression observed during interviews there is reason to believe that most Ss accepted the fact that the study had no connection with the prison authorities and Ss were not usually on "best behavior."

Picture-Frustration Study

The P-F was administered individually or in groups of no more than five Ss so that the testing performance could be readily observed. All the P-F protocols were scored according to Rosenzweig's Revised Scoring Manual (Rosenzweig, 1950) and without knowledge of the S's identity of his performance on other measures. The final scores yielded the percentage of responses falling within each type and direction of aggression provided for by the scoring system. Two scorers obtained 81% agreement on the exact scoring of responses.

In addition to analyzing protocols in terms of the usual P-F scoring system, responses scored as extra-punitive were rated along a scale designed to reflect the intensity of aggressive reaction to frustration (IE). Each response was assigned a rating according to the following three-point scale:

1. *Mild IE*: Responses which contain bland, indirect aggression within the limits of socially acceptable behavior, such as "I don't think so," or "I wish she would remove her hat."

2. *Moderate IE*: Responses which indicate sarcasm, strong displeasure, direct blame, or insistence that the other person do something about the situation, such as "How about next year," in re-sponse to an apology for breaking an appointment.

3. *Extreme IE*: Responses which contain direct, overt verbal aggression so that social prohibitions are weakened, as in violent expressions of emotion directed toward the frustrating person or situation. Swearing and other abusive language, including threats of physical harm such as, "I'll knock your head off," frequently occurred in such responses.

The IE scores referred to in the analysis of the data are median scores. The percentage of agreement on the exact rating of individual responses for two scorers was 83.8% and the coefficient of reliability (Pearson r) for the median scores was .86. Intropunitive responses were also rated along an intensity scale but the restricted range of expression precluded the development of a measure which could be utilized for further statistical analysis.

Aggression Measures²

In keeping with the conception of aggression as a heterogeneous pattern of behavior it seemed necessary to select measures designed to tap aggressive content in as broad a range of behavior as possible. Accordingly the P-F measures were related to data pertinent to aggressive behavior derived from the Rorschach, a psychiatric schedule, a specially devised Attitude Scale, and the prison record. The P-F was chosen because it seemed appropriate to evaluate the individual's manner of handling mild verbal frustration and provided a standardized system for the evaluation of aggressive responses to such situations. The psychiatric interview was intended largely as an exploration in the individual's personal history, including various aspects of his aggressive behavior as he saw it. The Attitude Scale was designed to permit the individual to express his attitudes about family members and his own feelings without such responses being filtered through an interpreter, such as an interviewer. The Rorschach was selected as an instrument presumed to minimize conscious control over the meaning of responses which might reflect aggressive feelings the S would not or could not express in response to the other techniques. The prison record provided some indication for the degree of antisocial behavior.

Rorschach

All Rorschachs were scored according to Klopfer's method, except for form level, where Beck's criteria were used. Records were evaluated with the help of a manual devised to analyze the following dimensions: degree and control of aggression, passive dependency, degree and control of sexual disturbance, degree of thought disturbance, principal defenses used, and diagnostic classification. Degree was rated from little (1) to extreme (3); control from good (1) to very poor (4). Signs for each dimension contained in the manual were obtained from the textbooks of Klopfer (1954), Phillips and Smith (1953), Rapaport (1945) and Schafer (1954), where two or more books agreed on their interpretation, as well as Elizur's (1949) hostility signs. The following excerpt of the description of the procedure given to the raters will illustrate the way in which these signs were used:

Very high frequencies of signs on a particular variable (e.g., aggression) will almost always indicate serious problems in that area, but a similar degree of concern may also be reflected in one or very few responses of extreme intensity. Thus many aggressive symbols such as knives and spears, or one response, in an otherwise bland record, of torn flesh and a lot of spattered blood, might both indicate a high degree of aggressivity. . . . Such evaluation requires much clinical experience, and the latter will be the principal tool of evaluation of the Rorschachs in this study.

² A more detailed description of these measures, including scoring criteria, may be obtained from the senior author.

It is recognized that this approach is quite subjective but the authors are not aware of any other method which can successfully deal with the Rorschach as a clinical instrument designed to tap largely unconscious feelings.

Space does not permit the listing of the signs used as indicators of all the dimensions, but the signs used in the evaluation of degree of aggressivity may serve as an example: Frequency of Elizur's (1949) hostility signs; blood; splattered fluids (e.g., paint); high visceral anatomy; fire; high *FM* with low *M* (example: rate "3" if more than 20% *FM* with no or one *M* where content aggression is moderate); very critical of cards; rejects (Card IX reject least important). As a reliability check, Rorschachs were rated according to the above method by another judge.³ Exact agreement varied from 40% to 75% on three-point scales, as poor and very poor control ratings were combined in order to roughly equalize *Ns* in all categories. Percentage agreements for the variables here reported are shown in Table 2. Two-step disagreements constitute 9.8% of all comparisons. Chi square significance at the .01 level for over-all comparisons and those reported here suggests better than chance agreement, but the percentages do not give much confidence for predictive purposes. The Rorschach ratings are nevertheless reported here because it was found that, with very few exceptions, Rorschach ratings made by the author correlated in the expected direction with other variables in the study (not reported here) so that there is some basis for believing that the criteria used are relevant to the categories of behavior examined. It is true that the Rorschach data does not add or detract much from the value of this report, but it seemed proper to report all the data pertinent to the variables under discussion.

Attitude Scale

Adopted from a procedure suggested by Toman (1955), a 205-item Thurstone-type scale was devised which required agreement or disagreement with statements expressing attitudes toward important objects and persons. The statements given to *Ss* were arranged in a random sequence and *Ss* were required to circle "Yes" or "No" for each statement. The scale was analyzed in terms of 45 five-item continua, ranging from maximum approval (Statement 1) to maximum disapproval (Statement 5) of a person or attitude. The following continuum of attitudes toward the father, for which data are reported in Table 2, will serve as an illustration of the instrument:

1. I loved my father best of all.
2. I liked my father.
3. I often liked my father, but sometimes I got a little fed up with him.
4. I often can't help having a feeling of dislike for my father.
5. It should be different, but I can't help hating my father.

³ Jacob Goldstein, New School for Social Research, New York City.

Agreement with Statements 1 or 2 or 1 and 2 or 1 and 3; or 2 and 3 or 1, 2, and 3 were scored as positive. Corresponding combinations of agreement to Statements 3, 4, and 5 were scored as negative. Any combination of agreements with statements on both sides of Statement 3 was scored as ambivalent, as was also agreement with Statement 3 alone. Nineteen percent of all responses were ambivalent. Response reliability was checked by a three-item lie scale (e.g., I have never had a headache) and by the repetition of three statements. Seventy-six percent of *Ss* disagreed with all three lie statements and 12% gave different responses on repeated statements.

Scale reliability was checked by independent sorting of all items by a naive rater into graded groups of five statements. The rater differed from a priori scaling in 3.6% of cases.

Psychiatric Schedule

The psychiatrist followed a prepared schedule which required that specific information be recorded in such a way as to be easily codable. All but 20 of the psychiatric interviews were conducted by one psychiatrist.⁴ All information was recorded during the interview. Coding criteria were specific enough to permit coding by clerical personnel. As a reliability check 1,000 randomly selected items, chosen from all psychiatric schedules, were recorded by an inexperienced coder. She achieved 83% exact agreement with the previous coder.

The following illustrates the criteria used in coding answers to questions about the frequency of aggressive behavior between *S* and the younger male sibling (see Table 2 for data).

- 0—no fights
- 1—infrequent fights, less than 10 instances in all
- 2—sometimes, but not more than once in three months
- 3—more frequently than above but not as much as once per week on the average
- 4—once per week or more, but less than once a day
- 5—once per day or more
- X—no such individual
- Y—no answer

Questions were not asked in a set way but the interviewer attempted to start each area of inquiry with a general question, such as, "How about your younger brother?" and then would ask more specific questions when necessary.

Prison Record

The last three items in Table 2 were taken from the prison record. History of serious antisocial aggression referred to previous felony convictions (assault). Present offense classification and the activity for which convicted (e.g., murder) were judged from a police description of the behavior during the offense. Thus the inactive accomplice to murder was

⁴ Hugh Storrow, now in the Department of Psychiatry, University of California, Los Angeles.

not classified either "aggressive" or "murderer" even though he may have been convicted for murder.

Twenty-two variables comprising all the data bearing on aggressive behavior of sufficient numerical range for statistical analysis were abstracted from these techniques and related to P-F performance by chi square (see Table 2). The distributions of E and IE scores were split into high, moderate, and low groups. Statistical requirements of chi square (Cochran, 1954) necessitated the collapsing of some contingency tables and in all such cases the upper extremes (high scores) were contrasted with the combined moderate and low scores.

RESULTS

Table 1 indicates that the mean extrapunitive, intropunitive, and impunitiveness scores for our sample closely matched the revised norms reported by Rosenzweig (1950). The scores for our sample, however, appear to be somewhat more variable on all three measures than those obtained from his normative sample. The mean for the intensity rating of extrapunitiveness was 1.29 with a standard deviation of .32 for a three-point scale, indicating that most Ss tend to express a rather mild intensity of aggression to this technique as measured by our ratings.

The interrelationships of P-F measures were determined by correlational and chi square analyses and are reported in Table 2. E was found to be significantly correlated with the IE measure, so that frequency and intensity of aggression appear positively associated. E was inversely related to intropunitive and impunitiveness as a result of the dependent scoring system. However, the association of both high E and high IE with both high ego defensiveness and low need persistence scores represents a relationship between independent scoring categories. The latter measures are two of Rosenzweig's three "types of Aggressive Reaction." According to Rosenzweig, Ss who express intense and/or frequent outward aggression on the P-F tended to respond in a manner "in which the ego of the subject predominated" rather than in a manner in which "the solution of the frustrating problem is emphasized" (Rosenzweig, Fleming, & Clarke, 1947, p. 166).

Twenty-three out of an over-all total of 73 relationships tested by chi square were significant at the .05 level or beyond, while 12 more approached significance at the .10 level. In relating E and IE to measures of

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR PICTURE
FRUSTRATION STUDY VARIABLES
(Rosenzweig's Norms in Parenthesis)

	Mean	Standard Deviation
Extrapunitiveness	47.49 (45)	15.71 (13.3)
Intropunitiveness	26.40 (28)	10.01 (8.3)
Impunitiveness	26.26 (27)	9.73 (9.5)
Intensity of Extrapunitiveness	1.29	.315

aggression elicited by other techniques, 12 of 22 relationships for E and 8 of 22 for IE were found to reach the .10 level or beyond. The total number of significant findings is thus well beyond the number expected on the basis of chance alone.

Chi squares significant at the .05 level or beyond revealed that E was associated with the S's report of aggression and negative feelings in object relationships, as with fathers and wives; the perception of peers as hostile and threatening; a history of antisocial aggression; and murder in the present offense. Moreover, the data reflect a number of trends which fall between the .05 and the .10 level. High E tends to be associated with the psychiatrist's rating of high verbal aggressiveness and, as judged from the Rorschach, poor control of aggression, the use of projection as a defense mechanism, and the diagnosis of schizophrenia. The only findings which seem inconsistent with the general direction of results was a trend for high E Ss to report infrequent fighting with peers.

The IE measure was found to correlate .48 with E—a moderate but statistically significant degree of correlation. It is not surprising, therefore, that IE related to many of the same measures as E. Thus, like high E, high IE relates to aggression and negative feelings in object relationships, as with siblings and wives; to a history of antisocial aggression; and to murder in the present offense. In addition, high IE was found to be significantly associated with poor control of aggressive impulses and the expression of intense affect on the Rorschach.

Both E and IE were unrelated to the degree or amount of aggression manifested on

TABLE 2

THE RELATIONSHIP OF EXTRAPUNITIVENESS (E) AND INTENSITY OF EXTRAPUNITIVENESS (IE) TO AGGRESSION VARIABLES

Variable Description	High E Scores are Related to:			High IE Scores are Related to:	
	df	Chi square ^a	Significance level ^b	Chi square ^a	Significance level ^b
P-F Variables					
High extrapunitiveness*	4				
High extrapunitiveness intensity	4	25.53	.01	25.53	.01
Low intropunitiveness	4	65.35	.01		
Low impunitiveness	4	31.90	.01	12.59	.02
Low object dominance	4	7.80	.10	10.32	.04
High ego defensiveness	4	37.10	.01	5.48	
Low need persistence	4	24.97	.01	16.54	.01
				10.46	.04
Rorschach					
Degree of aggression (73%) ^c	2	1.87			
Inadequate control of aggression (65%)	2	5.17	.07	1.24	
High degree of intense affect (62%)	4	2.71		6.63	.04
Use of projection as a defense (73%)	1	2.89	.10	10.23	.05
Diagnosis of schizophrenia (65%)	1	4.46	.04		
Attitude Scale					
Reaction to frustration	2	.98			
Reaction to criticism	2	2.04		1.59	
Perception of peers as hostile	2	6.89	.04	.32	
Negative feeling toward father	2	14.81	.01	2.85	
				3.86	
Psychiatric Schedule					
Subject's rating of self-control over anger	1	.87		.49	
Psychiatrist's rating of subject's aggressiveness	2	.58		1.31	
Tendency to express aggression verbally (Psychiatrist's rating of manner in which subject expresses aggression)	1	3.24	.08	1.23	
Submissive behavior toward father (developmental)	2	2.92		11.37	.01
Frequent aggressive behavior toward younger male siblings	1	1.35		5.47	.02
Infrequent fighting with peers (developmental)	1	3.56	.07	.59	
Frequent destruction of property (developmental)	1	3.28		3.47	.07
Fighting with peers (current)	1	.31		.64	
Frequent verbal aggression toward wife	1	2.78	.10	1.02	
Frequent physical aggression toward wife	2	6.97	.03	5.42	.07
History of serious anti-social aggression	2	8.40	.01	3.18 ^d	.08
Present offense classified as sexual, aggressive, or acquisitive	4	1.62		4.11	
Murder in present offense	2	4.68	.10	6.83	.04

^a Yates correction in two by two tables.

^b Two-tailed.

^c Exact agreement of two raters on three-point scale.

^d df = 1.

* $r = .48$ significant at the .01 level.

the Rorschach, S's report of his reaction to frustration and criticism, S's rating of his self-control over anger, interviewer's rating of S's aggressivity, current fighting with peers and present offense classified as sexual, aggressive, or acquisitive.

Both amount and intensity of extrapunitive-ness were compared with age, race, religion, rural-urban background, social class, marital status, education, IQ, and length of time in prison. Those Ss with high E tend to be older ($\chi^2 = 9.86$, $df = 2$, $p = .01$), widowed or divorced ($\chi^2 = 3.71$, $df = 1$, $p = .06$), and to have an education beyond the 8th grade ($\chi^2 = 7.17$, $df = 2$, $p = .04$). Ss with high IE tend to come from an urban background ($\chi^2 = 7.56$, $df = 2$, $p = .03$). None of the other comparisons yielded chi squares approaching significance.

The relationships of intropunitiveness and impunitiveness to other measures discussed above were investigated but results have not been presented since they were generally found to be the inverse of those relationships reported for extrapunitive-ness by virtue of the reciprocal nature of the scoring system.

DISCUSSION

This investigation has provided evidence that the P-F relates to a variety of manifestations of aggressive behavior as assessed by a psychiatric interview, Rorschach evaluation, and an attitude scale. The scoring dimension of E related to aggression in terms of hostile, aggressive interpersonal feelings and behavior, and antisocial aggression with tendencies toward the use of projection and inadequate impulse control. The score for IE developed in this study and the standard scoring dimension of extrapunitive-ness (E) show considerable overlap, the former also relating to the adequacy of impulse control and intensity of affect on the Rorschach. Determination of the value of the intensity measure in extending the range of the P-F scoring system is a problem for further research.

The results also indicate that higher education and IQ are positively related to high E and high IE. An additional analysis of the data revealed that IQ and education were not significantly correlated with other variables bearing on aggression so that no attempt was

made to control their effects statistically. One could speculate that the P-F technique was most suitable for eliciting aggressive responses from Ss in our sample who are better able to deal with verbal stimuli or to express aggression verbally.

The P-F failed to relate to many aspects of aggression, such as type of criminal offense, reported behavior in response to frustration and criticism, and aggressive behavior toward peers. These measures which did not relate to the P-F seem no less a reasonable criteria than those which had a positive relationship to it. The problem of evaluating validity in terms of several relationships, some of which are significant or nearly significant while others are not, may be viewed in at least two ways. One is to postulate that items which relate positively to the P-F should characterize some particular aspect of aggression (e.g., level). Examination of Table 2 indicates that for our results such a conceptualization would, at best, require considerable and perhaps tenuous speculation, inasmuch as diverse features of aggressive behavior relate to extrapunitive-ness. This failure to find a cluster of conceptually meaningful relationships may be the fault of the study, but we would suggest that expecting such a cluster may not be reasonable, which brings us to the alternative interpretation.

Suppose, for instance, that there are two Ss with the same high E P-F scores for whom extensive personality evaluations, presumed valid, are available. One of these Ss generally expresses his aggression only verbally. He is sarcastic, bitter, and superior but never attacks anyone directly and has no history of overt physical aggressive behavior. For the other individual the high E score is but one manifestation of generally very aggressive behavior expressed along the entire spectrum of this attribute. There would probably be none, or a negative relationship between the P-F and measures of overt aggression for the first S and a positive relationship with almost any measure of aggression for the second S. This illustrates how the same form and intensity of aggression in a particular situation (e.g., high E on the P-F test) may reflect different types and degrees of aggressivity for different individuals. To be sure, aggression shown on

one technique, such as the P-F, must be reflected in *some* other measures of that attribute if the technique is valid; however, *which* measures will be found to correlate would, from this point of view, be a problem of sampling. That is, positive correlations can be expected to the degree to which the personality characteristics of different Ss overlap in the expression of aggression. Since even elaborate matching procedures are unlikely to yield identical samples of Ss, different samples can be expected to show different interrelationships between a particular measure, such as the P-F, and other measures designed to reflect a range of that attribute (e.g., Rorschach, attitude scale, other measures of aggression).

Nevertheless, while two samples would rarely yield exactly the same interrelationships, it is possible that given a large number of successive samples tested with the same battery, the test measure (e.g., the P-F) will be found to correlate more often with some criterion measures than with others. Thus, while the pattern of intercorrelations may not make too much sense for any particular sample (as in the present study), it is possible that a conceptually meaningful pattern of test-criteria relations may emerge through frequency counts of such relations across a large number of samples. Such a procedure might help to establish the specific construct validity of the P-F by indicating which aspects of aggression it reflects more often than others. The present study is concerned with the much more modest task of investigating whether the P-F relates to *any* aspect of aggression. For this purpose it would seem a reasonable test of validity to require that, given a pool of criterion measures (e.g., attitude scale, Rorschach, etc.), the number of positive relations to the measure tested (P-F) exceed chance expectations.

SUMMARY

Extrapunitiveness in the Picture-Frustration Study and an intensity of extrapunitiveness measure developed by the authors were related to 22 measures of aggression derived from the Rorschach, an attitude scale, a psychiatric interview schedule, and case history data. The Ss in this study were 121 inmates

of a state prison selected so as to match the entire prison population. Of a total of 44 chi square comparisons between the Picture-Frustration Study and these measures, 10 were significant at better than the .05 level of confidence, and another 10 between the .05 and .10 level. These results were interpreted as indicating that the P-F has some relation to other measures of aggression. The P-F was not found to tap any particular level or aspect of aggression, and it was suggested that repeated cross-validation would be required before possible criteria validity could be established.

REFERENCES

- AMERICAN PSYCHOLOGICAL ASSOCIATION, Committee on Test Standards. *Technical recommendations for psychological tests and diagnostic techniques*. Washington: APA, 1954.
- COCHRAN, W. G. Some methods for strengthening the common chi-square test. *Biometrics*, 1954, 10, 417-451.
- ELIZUR, A. Content analysis of the Rorschach with regard to anxiety and hostility. *J. proj. Tech.*, 1949, 13, 274-284.
- FRY, F. D. A study of reactions to frustration in 236 college students and in 207 inmates of state prisons. *J. Psychol.*, 1949, 28, 427-438.
- HOLZBERG, J. D., & HAHN, F. The picture-frustration technique as a measure of hostility and guilt reaction in adolescent psychopaths. *Amer. J. Orthopsychiat.*, 1952, 22, 776-795.
- KLOPPER, B. *Developments in the Rorschach technique*. New York: World Book, 1954-56.
- PHILLIPS, L., & SMITH, J. G. *Rorschach interpretation*. New York: Grune & Stratton, 1953.
- RAPAPORT, D. *Diagnostic psychological testing*. Chicago: Year Book, 1945-1946.
- ROSENZWIEG, S. Revised norms for the adult form of the Rosenzweig Picture-Frustration Study. *J. Pers.*, 1950, 18, 344-346.
- ROSENZWIEG, S., FLEMING, E. E., & CLARKE, H. J. Revised scoring manual for the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1947, 24, 165-208.
- SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954.
- TOMAN, W. The Multiple Attitude Test: A diagnostic device. *J. abnorm. soc. Psychol.*, 1955, 51, 163-170.
- VANE, J. R. Implications of the performance of delinquent girls on the Rosenzweig Picture-Frustration Study. *J. consult. Psychol.*, 1954, 18, 414.
- WEINBERG, W. L. A study of the relationship of the extrapunitive category of the Rosenzweig P-F Study to overt behavioral aggression in prisoners. Unpublished master's thesis, Univer. Oregon, 1952.

(Received October 12, 1959)

A FACTOR ANALYSIS OF GOUGH'S CALIFORNIA PSYCHOLOGICAL INVENTORY¹

JAMES V. MITCHELL, JR. AND JOHN PIERCE-JONES

University of Texas

The California Psychological Inventory (CPI) developed by Gough (1957) during the past decade is a structured verbal inventory designed to measure such personality characteristics as dominance, tolerance, responsibility, etc. These characteristics are thought to be more important for comprehending normal behavior than are the traits sampled by instruments like the MMPI, which appear to be most useful in describing psychopathological conditions and maladjustment syndromes (cf. Sloan & Pierce-Jones, 1958). The 18 scales of the CPI are designed to yield a presumably meaningful set of scores which can provide a profile representing the personality pattern of an individual. Shaffer (1959) has written that "the CPI appears to be a major achievement." Cronbach (1959), while eschewing Gough's frankly atheoretical approach to personality measurement, has applauded Gough's apparent technical psychometric skill. Thorndike (1959), however, has asserted that the scales of the CPI "... provide a very redundant, inefficient, and confused picture of individual personalities. . . . Of the 18 scales, there are only 4 that fail to correlate at least .50 with some other scale."

PROBLEM

The present investigation was motivated, in part, by our difficulty in comprehending individual CPI profiles, a difficulty resulting from the conditions so clearly indicated by Thorndike's comments mentioned above. We were also puzzled by Gough's grouping of

¹ This investigation was conducted with support from the Mental Health in Teacher Education Project at the University of Texas, a project which is supported by a grant from the National Institute of Mental Health.

the 18 CPI scales into several named clusters shown in the *Manual* (Gough, 1957), since there appeared no clear empirical basis for such clusters in the correlation matrices provided. In consequence, the factor analytic research reported in this paper was undertaken to obtain the kind of evidence that would help to shed light on the empirical justification of the scales and scale groupings offered by Gough.

PROCEDURE

A total of 258 cases was employed in this investigation. This sample included 213 females and 45 males who were enrolled for study in a teacher training curriculum. The subjects (Ss) generally had major fields of academic work in university subjects other than Education.

A correlation matrix² was constructed which consisted of the 153 CPI scale intercorrelations which had been computed by the product-moment method utilizing an IBM Model 650 electronic computer. From this matrix four factors were extracted by means of the centroid method (Guilford, 1954). Application of Humphrey's rule (Fruchter, 1954) and Burt's empirical formula (Thomson, 1951) indicated that factorization should be terminated after extraction of the fourth factor. Orthogonal rotations were then performed by means of Kaiser's (1958) Varimax technique, the characteristics of which have been discussed in recent papers by Comrey (1957a, 1957b, 1957c).

RESULTS

The rotated factor matrix is shown in Table 1. The four factors accounted for 26%, 15%, 7%, and 12% of the total variance,

² A one-page table of the CPI scale intercorrelations used in this study has been deposited with the American Documentation Institute. Order Document No. 6329 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 1
ROTATED FACTOR MATRIX FOR THE CALIFORNIA PSYCHOLOGICAL INVENTORY

CPI Scales	I	II	III	IV	<i>h</i> ²
Class I Measures of Poise, Ascendancy, and Self-Assurance					
1. Dominance	18	76	04	-09	61
2. Capacity for Status	21	59	04	50	64
3. Sociability	27	78	04	15	71
4. Social Presence	-01	62	-23	51	70
5. Self-Acceptance	-17	77	08	05	63
6. Sense of Well-Being	79	16	05	23	70
Class II Measures of Socialization, Maturity, and Responsibility					
7. Responsibility	58	09	44	21	58
8. Socialization	43	02	57	-06	51
9. Self-Control	92	-19	08	-01	77
10. Tolerance	67	11	15	54	78
11. Good Impression	83	08	-08	-06	71
12. Communality	02	06	58	-02	34
Class III Measures of Achievement Potential and Intellectual Efficiency					
13. Achievement by Conformance	80	25	23	08	77
14. Achievement by Independence	47	02	11	67	69
15. Intellectual Efficiency	46	24	16	53	58
Class IV Measures of Intellectual and Interest Modes					
16. Psychological Mindedness	47	22	-16	32	40
17. Flexibility	-12	02	-25	56	39
18. Femininity	-02	-02	45	-04	21
Percentage of Total Variance	26%	15%	7%	12%	

Note.—Decimal points omitted before all factor loadings in table.

respectively. Analysis of the factor loadings resulted in the following factor descriptions.

Factor I, probably the most important factor of the present analysis, had its highest loadings for the CPI scales named Self Control, Good Impression, Achievement via Conformance, Sense of Well Being, Tolerance, and Responsibility. Provisionally, it appears that this factor might well be named *Adjustment by Social Conformity*. Factor II had five relatively high loadings, ranging from .59 to .78. The CPI scales named Dominance, Capacity for Status, Sociability, Social Presence, and Self-Acceptance were the important ones involved in this factor, suggesting that it should be named *Social Poise* or, alternatively, *Extroversion*. Factor III was not as well defined as the previous factors, but it had loadings above .40 for the CPI scales

labeled Responsibility, Communality, Socialization, and Femininity. Gough's (1957) original descriptions of characteristics associated with these scales strongly suggest the serious, responsible, conscientious person, so we have tentatively chosen to name this factor *Super-Ego Strength*. Factor IV is possibly the most interesting of the factors identified in the present analysis. Having loadings of .50 or higher for CPI scales called Tolerance, Intellectual Efficiency, Capacity for Status, Flexibility, Social Presence, and Achievement via Independence, this factor suggests a complex of qualities which might augur well for success in a wide range of human activities. Common to all of these qualities is an emphasis on intellectuality, broad interests and perspectives, and thoroughgoing independence. Our present disposition is to name this

factor *Capacity for Independent Thought and Action*.

DISCUSSION

Gough (1957) has classified the 18 CPI scales into the following four groups:

- Class I: Measures of Poise, Ascendancy, and Self-Assurance
- Class II: Measures of Socialization, Maturity, and Responsibility
- Class III: Measures of Achievement Potential and Intellectual Efficiency
- Class IV: Measures of Intellectual and Interest Modes

The factor analytic results reported in this paper lend support to four CPI scale classes of somewhat different composition from Gough's, as can be inferred from the results and factor descriptions already presented in Table 1. It is interesting to notice, for example, that our Factor II, Social Poise or Extroversion, encompassed five of the six scales included in Gough's Class I. If Gough's four classes of scales are viewed as hypotheses concerning factor structure, his first class appears from our analysis to be a reasonable one, except that we would be forced to exclude the Well-Being scale from the class. Our Factor III, here called Super-Ego Strength, identified three of Gough's Class II measures and one scale, Femininity, from Gough's Class IV. Factor I identified in the present analysis located five Class II scales, three Class III scales, and one scale from each of Classes I and IV when loadings of .43 or higher were considered. Our Factor IV, Capacity for Independent Thought and Action, had two substantial loadings for scales included by Gough in his Class I, one loading representing a Class II measure, two loadings for scales included in Class III, and one loading representing a measure included by Gough in his Class IV. If the results obtained in the present analysis are valid, it seems clear that Gough's Class IV is unacceptable as a cluster of related scales, since its constituent scales load on three separate factors. Similarly, Class III as defined by Gough does not seem to be empirically justified. Finally, Gough's Class II appears to be a mixture of scales which, in the present analysis, have their loadings on Factors I

and III. If the CPI continues to be scored for 18 scales, and if a defensible means for grouping the scales into classes is desired, classification should probably be based upon factor analytic findings such as those presented in this report.

It seems quite apparent from the results of this research that the CPI cannot be regarded with real justification as measuring the 18 relatively independent personality dimensions that it is purported to measure. It is also true, judging by our results, that individual personality profiles might well be based on only a few selected CPI scales. Perhaps, for example, the Self-Control scale can be regarded as virtually a pure measure of Factor I, Adjustment by Social Conformity, since it loaded .92 on that factor and was essentially independent of the three other factors. The Dominance, Sociability, and Self-Acceptance scales seem equally good measures of Factor II, Social Poise, and they lead to no appreciable inferences regarding Factors I, III, and IV. Factor III, Super-Ego Strength, is well estimated by the Communality and Femininity scales, which appear to be independent of the three remaining factors. The Flexibility scale occupies a similar position with regard to Factor IV, Capacity for Independent Thought and Action. In each case the decision to use a scale or scales to represent a factor would naturally be contingent upon reliability data as well as factorial composition.

It is interesting to note that with the exception of the Communality and Sociability scales, those scales which seem best to represent the four factors identified in the present analysis lead to personality descriptions couched in long-used psychological terms. In this regard, it can be noted that Cronbach (1959) has argued for such descriptions over others couched in terms like "social presence," which apparently refer to social behavior patterns often regarded as complicated resultants of motivational dispositions, abilities and skills, and situation-linked stimulus conditions.

While the present analysis may have provided some clarification of personality measurement using the CPI, it should be acknowledged that one limitation of this study exists

because factor identification and description are partially dependent upon Gough's original selections of items for scales and upon the verbal descriptions assigned to the scales. But this fact in no way negates our earlier conclusion that the 18 CPI scales represent a much smaller number of personality dimensions. Nor, indeed, does the fact that our factor interpretations made use of the original scale descriptions lessen the significance of these interpretations, since a careful review of the items included in the factors tends to support the designations which we have assigned to each.

SUMMARY

A centroid factor analysis was carried out with the 18 scales of the California Psychological Inventory. The sample employed consisted of 258 students enrolled in an introductory course in educational psychology at the University of Texas. Four factors were extracted and rotated by Kaiser's Varimax method. The factors identified were tentatively named: I. Adjustment by Social Conformity, II. Social Poise or Extroversion, III. Super-Ego Strength, and IV. Capacity for Independent Thought and Action. It was suggested that individual personality profiles based on the CPI might use only those scales best representing the four factors identified in this research. Such a limitation would produce profiles permitting personality descriptions to be made in such conventional psychological terms as "dominance," "self-acceptance," and the like, rather than in such

complex social behavioral terms as, for example, "social presence" and "capacity for status."

REFERENCES

- COMREY, A. L. A factor analysis of the items on the MMPI hypochondriasis scale. *Educ. psychol. Measmt.*, 1957, 17, 568-577. (a)
- COMREY, A. L. A factor analysis of the items on the MMPI depression scale. *Educ. psychol. Measmt.*, 1957, 17, 578-585. (b)
- COMREY, A. L. A factor analysis of the items on the MMPI hysteria scale. *Educ. psychol. Measmt.*, 1957, 17, 586-592. (c)
- CRONBACH, L. J. California Psychological Inventory. In O. K. Buros (Ed.), *The fifth mental measurements yearbook*. Highland Park, N. J.: Gryphon, 1959.
- FRUCHTER, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.
- GOUGH, H. G. *Manual for the California Psychological Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1957.
- GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.
- SHAFFER, L. F. California Psychological Inventory. In O. K. Buros (Ed.), *The fifth mental measurements yearbook*. Highland Park, N. J.: Gryphon, 1959.
- SLOAN, T. J., & PIERCE-JONES, J. The Bordin-Pepinsky diagnostic categories: Counselor agreement and MMPI comparisons. *J. counsel. Psychol.*, 1958, 5, 189-195.
- THOMSON, G. H. *The factorial analysis of human ability*. Boston: Houghton Mifflin, 1951.
- THORNDIKE, R. L. California Psychological Inventory. In O. K. Buros (Ed.), *The fifth mental measurements yearbook*. Highland Park, N. J.: Gryphon, 1959.

(Received October 22, 1959)

THE DEVELOPMENT OF AN AFFECT ADJECTIVE CHECK LIST FOR THE MEASUREMENT OF ANXIETY

MARVIN ZUCKERMAN¹

Institute of Psychiatric Research, Indianapolis

Most personality tests are designed to measure relatively stable traits. Testees respond to items asking how they "generally," "often," "usually," "seldom," etc. behave. The time referent is vague and subjects (Ss) may interpret the questions as referring to the last week, month, year, or their entire lifetime. Many of the traits in which clinicians are interested, particularly affective traits such as hostility or anxiety, are assumed to show large intra-individual variation from hour to hour and from day to day. While the concept of a "general level of anxiety" may be useful for gross discrimination, there are many occasions where one would like to measure changes in anxiety over shorter periods of time. Experiments where an attempt is made to induce anxiety, or experiments on the effect of the "tranquilizer" drugs are examples of research where temporal ambiguity of the usual anxiety questionnaires might make them insensitive to change. Experiments of this kind conducted at the Institute of Psychiatric Research led the author to the development of a test which could be given quickly, scored objectively, and adapted for varying time sets.²

The adjective check list method seemed ideally suited for the purposes discussed above. Gough (1955) has used a check list test for developing empirical scales of various personality traits. Nowlis (1953) reported the use of a check list in measuring changes induced by drugs given to college students.

The purpose of the present study was to develop empirically a scoring key for "anxiety" using a pool of adjectives with various

affective connotations, and to test the reliability and validity of this anxiety score.

EMPIRICAL DEVELOPMENT OF A SCORING KEY

Adjectives with affective connotations were collected from Gough's and Nowlis' lists and from a thesaurus. Adjectives which were of low frequency in the written language were excluded so that Ss of less than average intelligence could understand the items. The final list of adjectives consists of 61 items.³

The scoring key was derived from item analyses in two studies. The first study by Persky, Maroc, Conrad, and den Breeijen (1959) compared a group of psychiatric patients rated high in anxiety with a group of normal controls rated low in anxiety on the basis of a psychiatric interview. A preliminary scoring key of 30 words was used, based on an a priori selection of words with "anxiety" connotations. This score yielded a highly significant difference between the groups. An item analysis was performed by comparing the frequencies in each group checking, or not checking, each of the adjectives. Twenty-four of the words yielded significant differences ($p < .05$). Twelve of these were "anxiety-plus" words (checked more frequently by the High Anxiety group) and 12 were "anxiety-minus" words (checked more frequently by the Low Anxiety group). Although the two groups in this study were not ideally matched on nonanxiety variables it is interesting that all but one of the anxiety-plus words had been previously selected as anxiety words on the basis of their connotation and were among those used in the a priori scoring.

The second study by Levitt, den Breeijen, and Persky (in press) measured the effects of

¹ Now at Department of Psychology, Brooklyn College.

² The idea for this test emerged out of a discussion with John I. Nurnberger on this problem.

³ A copy of the check list may be secured by writing to the author.

a hypnotically induced anxiety state in normals. A score on the Affect Adjective Check List (AACL) based on the item analysis in the preceding study showed a highly significant rise during the anxiety condition. An item analysis yielded 38 words showing significant changes in checking frequency during the anxiety condition. Twenty-one were anxiety-plus words (increase in checking frequency during anxiety condition) and 17 were anxiety-minus words (decrease in checking frequency during anxiety condition).

The final scoring key includes 21 words which proved to be significantly related to anxiety in *both* of the aforementioned studies. These words are listed below.

Anxiety-plus: afraid, desperate, fearful, frightened, nervous, panicky, shaky, tense, terrified, upset, worrying.

Anxiety-minus: calm, cheerful, contented, happy, joyful, loving, pleasant, secure, steady, thoughtful.

Anxiety-plus words are scored 1 if checked, and anxiety-minus words are scored 1 if not checked. The possible range of scores is 0 to 21.

RELIABILITY STUDY: COLLEGE GROUP I

Two versions of the AACL and the Taylor (1953) Manifest Anxiety Scale (MAS) were given to 50 students in two sections of an elementary psychology class at the Purdue University Extension School. The group consisted of 43 males and 7 females. The average age was 22.3 ($SD = 4.6$). The items in the two versions of the AACL were the same but one test asked the Ss to check words which described how they "generally feel" while the other test asked them to check items describing how they felt "today." On the latter test they were told that "today" was defined as beginning from the time they awoke that morning. One section received the General test first and the Today test second while the other section took them in the reversed order. One week later the Ss took the two AACLs again and each section received them in an order reversed from the first session.

Tests for the effect of order on the General and Today tests were performed by comparing each S's scores on the occasions when he took the tests first and when he took them second. The 50 Ss scored an average of .68

point higher on the Today test when it was taken before and .32 point higher on the General test when it followed the Today test, but these differences were not significant (t 's = 1.01, .95). One can conclude that order had no effect on the two AACL versions.

Two kinds of reliability were examined for the General and Today tests: internal consistency on the first testing, and retest reliability from the first to the second test. The General test was expected to show reasonably high internal reliability and retest reliability since it was assumed that the General time set would cause Ss to describe a stable trait. The Today test was expected to show high internal reliability on a single testing but low reliability on retest. Since the Today test is designed to measure day to day fluctuations it is unlikely that it could be sensitive to these fluctuations and still remain stable from week to week.

The two measures of reliability for each of the tests is given in Table 1. Internal reliability is calculated using the Kuder-Richardson Formula 20. It can be seen that the expectations of the experimenter (E) are borne out. The General test is reliable internally and in retest, while the Today test is internally reliable on a single testing but low in reliability on a retest. This contrast in reliabilities bears out the E 's assumptions about the different nature of tests attempting to measure stable and fluctuating traits. The actual correlation between the two versions of the AACL on the first testing was .43, indicating a significant, but only moderate, relationship between the two tests.

VALIDITY STUDY: COLLEGE GROUP II

The purpose of this study was to see if the Today AACL anxiety score would show an in-

TABLE 1
RELIABILITIES OF GENERAL AND TODAY
VERSIONS OF AACL

	K-R 20 ^a	Retest
General	.72**	.68**
Today	.85**	.31*

^a Significance of K-R 20 was established by the formula:
 $F = 1/(1 - r)$, $df = N - 1/N - 1 \times \text{No. items} - 1$.

* Significant below .05 level.

** Significant below .001 level.

crease when this test is given on the day of an examination. "Examination Anxiety" is not usually as intense as the anxiety seen in clinical patients and it varies considerably between individuals. However, if the AACL proved sensitive to Examination Anxiety it would probably be sensitive to more intense forms of anxiety.

Method

The Ss were 35 college students in a section of elementary psychology. Twelve were females and 23 were males. The average age was 19.5 ($SD = 2.7$). One student dropped the course and his data was only used on comparisons involving the initial tests.

The students were given the General and Today versions of the AACL, the MAS, and the Barron Ego-Strength scales (Barron, 1953) on the second day of the class. They were given the Today AACL at the beginning of every other class meeting except those following test day meetings. The *E* was interested in comparing the nonexam day AACL scores with the exam day AACL scores. The AACL was given regularly at the same time so that the students would not guess the connection between the AACL and their examinations. At the end of the course they were asked if any of them had guessed the connection and none claimed to have seen the purpose of the experiment. The AACL was not given at the three meetings after examinations because at these times the students were waiting to hear their grades and the *E* was not sure whether these occasions should also be classified as "anxiety" situations. There were a total of 10 nonexam days and 3 exam days. All 34 students were present on the exam days. Absences varied on the other days from one to three students. No student was absent more than once on the 13 days the AACL was given.

Results

Figure 1 shows the change in AACL anxiety scores during the course of the 13 class meetings. In the case of absent Ss, the mean score on the other 9 nonexam days was sub-

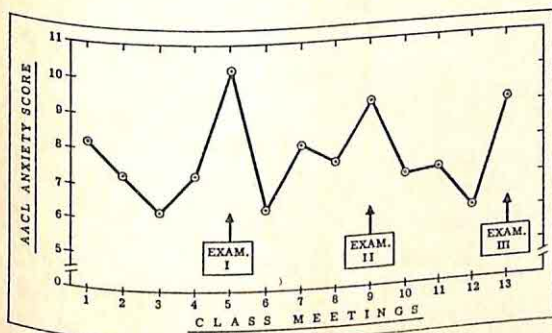


FIG. 1. Changes in AACL anxiety scores during 13 class meetings.

TABLE 2
MEAN DIFFERENCES BETWEEN EXAM AND
NONEXAM DAY ANXIETY SCORES

	M_D (exam- M non- exam ^a)	S_{MD}	t	p^b
Exam I	+2.91	.54	5.39	<.0005
Exam II	+2.13	.76	2.80	<.005
Exam III	+2.49	.69	3.61	<.001
M all exam days minus M all nonexam days	+2.41	.53	4.55	<.0005

^a Mean on the three or four nonexam days subsequent to the last exam and preceding the current exam, excluding days when the results of exams are reported.

^b One-tailed test.

stituted for the missing entries in order to calculate comparable group means for the different sessions. The means of the group on nonexam days ranged from 6 to 8. The SD of the 10 nonexam day means was .77. On the three exam days the AACL scores rose to 9 or 10. Table 2 gives the mean differences between: (a) AACL scores on each of the exam days and the means of these scores on the three or four nonexam days immediately preceding each exam, (b) the mean anxiety score on all three exam days and the mean score on all 10 nonexam days. The AACL increases are significant for each of the exam days and the mean for all exam days is significantly higher than the mean for all nonexam days. The hypothesis that pre-examination anxiety will result in increased anxiety scores on the AACL is supported by these results.

After these results had been obtained, another hypothesis occurred to the *E*. It is possible that the amount of anxiety elicited by an examination might be related to how well one expects to do on the examination. Although no measures had been made of the Ss' anticipations prior to the examination, their actual performance on the examination might be some measure of their preparation and consequent anticipation. The class was divided into thirds on the basis of their grades on each of the examinations. The mean changes on the AACL anxiety score of the High and Low Grade groups were then compared. The results of these tests are contained in Table 3. It can be seen that on all three examinations the Ss who made high grades

TABLE 3

COMPARISONS OF MEAN DIFFERENCES BETWEEN EXAM
AND NONEXAM DAY ANXIETY SCORES IN HIGH
AND LOW GRADE GROUPS

	High Grade		Low Grade		<i>t</i>	<i>p</i> ^b
	<i>N</i> ^a		<i>N</i>			
Exam I	11	+1.57	11	+4.55	2.32	<.025
Exam II	12	+0.10	11	+3.59	2.20	<.025
Exam III	11	+1.45	11	+3.76	1.47	<i>ns</i>
<i>M</i> all exam days minus <i>M</i> all nonexam days	11	+0.75	12	+3.21	2.02	<.05

^a Differences in *N* are because of ties in grade scores at the cut-off point.

^b One-tailed test.

showed less of an increase in anxiety scores than the *Ss* who made low grades. The differences are significant for the first two examinations but not significant for the third examination. The class was then divided into thirds on the basis of their total percentage score for the entire term. The High and Low Grade students were compared on the differences between the mean for all exam days and the mean for all nonexam days. This result is also contained in Table 3. The difference is significant at the .05 level. The correlation between the total percentage grade for the term and the mean exam minus nonexam day differences for the total group of 34 was $-.34$, which was significant below the .05 level. The results generally support the hypothesis that the students doing well on examinations would show less increase in anxiety scores than the students doing poorly on examinations. However, these results do not definitively support the idea that the student's anticipation of success or failure is responsible for this difference. An alternate hypothesis might be that the pre-examination anxiety actually caused the poorer performance on the examination. In the next study of this question, the *E* hopes to obtain anticipations before the examination, on the hypothesis that these will be more strongly related to the anxiety scores than the actual grades on the examinations.

VALIDITY: RELATIONSHIPS WITH OTHER TESTS—GROUPS I AND II, AND PREGNANT WOMEN GROUP

Since it has been hypothesized that most questionnaires give the *S* a general time set it

was expected that the MAS and other tests would correlate significantly with the General version of the AACL but not with the Today version given on one particular day. However, it was expected that the MAS would correlate significantly with the mean of a number of AACL anxiety measures since this would presumably be a better measure of the general anxiety level than a score on a single day. The MAS should correlate significantly with the Month version of the AACL given to the pregnant women since a month is a long enough period to consider a general time set.

Method

In each of the preceding college groups the MAS was given to the *Ss*. In College Group II the *Ss* also took the Barron Ego-Strength scale. A third group of *Ss* was collected in another study on pregnancy.⁴ This study involved monthly interviews and tests with a group of 51 pregnant women. The women were mainly from a lower-class population. The average age was 21.1 ($SD = 5.2$). The average education in years was 11.3 ($SD = 2.1$). The group consisted of 41 Negro and 11 white *Ss*. The AACL used in this study asked the *Ss* how they had been feeling during the past month. At the time of the first interview, the *Ss* were given the Month AACL, MAS, the MMPI M-F scale, and three scales from the Parental Attitude Research Instrument (Schaefer & Bell, 1958), the latter constituting a factor called "Hostility-Rejection" (Zuckerman, Ribback, Monashkin, & Norton, 1958). At each subsequent interview the *Ss* were again given the AACL and the MAS. The number of interviews ranged from 1 to 6. A subgroup of 10 *Ss* had taken the AACL and the MAS during the sixth, seventh, eighth, and ninth months of pregnancy.

Results

The correlations between the MAS and the various versions of the AACL given to the three groups of *Ss* are presented in Table 4. In the College Group I the MAS correlated significantly with the Today AACL ($r = .29$) and did not correlate significantly with the General AACL ($r = .18$). The direction of the difference in this group was contrary to expectation but the difference between these correlations was not significant ($t = .78$). In the College Group II the MAS correlated significantly with the General ($r = .58$) and did not correlate significantly with the Today AACL ($r = .32$), according to expectation.

⁴ This study is still in the data analysis stage.

The correlation between the MAS and General AACL was significantly higher than the correlation between the MAS and Today AACL ($t = 2.07$, $p < .05$). The differences in results between these two similar samples are difficult to explain. The MAS correlated significantly with the mean of 10 Today AACL scores ($r = .52$), as expected. It did not correlate significantly with the mean of the AACL scores on the three examination days ($r = .28$). The difference between these latter two correlations was not significant ($t = .63$). However, the correlation between MAS and AACL on the first examination day was .40 which was significant below the .05 level. The correlation with AACL on the second exam day dropped to .09. The correlation for the third exam day was .19. Both of the latter correlations were not significant. The differences between the correlation on the first exam day and the correlations on the second and third exam days were not significant (t 's = 1.75, 1.26).

The correlations of MAS with AACL for the group of pregnant women are also contained in Table 4. The correlation between the tests on the first occasion when they were both given ($N = 51$) was .65, significant below the .001 level. The more General time set of a Month resulted in a significant correlation with MAS, as expected. The relationship

with MAS as a function of the particular month of pregnancy was measured in the 10 Ss who had taken both tests during the sixth to ninth months. The average MAS scores showed a significant drop during this period while the AACL anxiety scores showed no trend. It can be seen that the correlations between the two tests are very high during the sixth, seventh, and eighth months but drop to a nonsignificant .38 during the ninth month. The difference between the correlation of the two tests in the eighth month and in the ninth month was not significant ($t = 1.77$).

In the College Group II, the Barron Ego-Strength scale correlated significantly with the General AACL score ($r = -.37$), and nonsignificantly with the mean for exam days ($-.28$) and the mean for nonexam days ($-.34$). The mean AACL scores in the Pregnant Group did not correlate significantly with the parental attitude scales (Marital Conflict, $r = .14$; Irritability, $r = .02$; Rejection of the Homemaking Role, $r = .21$), and it did not correlate significantly with the *M-F* scale ($-.14$), or with a score based on physical complaints in the interview (.20). The failure of the AACL to correlate with these measures is not crucial for the validity of the test, for these tests are not intended as direct measures of anxiety.

AACL AND SEX, AGE, AND EDUCATION

The total college group used in the first two studies consisted of 65 males and 19 females. The differences between the means for males and females on the General and Today versions of the AACL were compared and found to be small and insignificant (General, $t = .09$; Today, $t = .49$). In the same group of 84 Ss, age was not found to correlate with the General AACL ($r = .08$) or with the Today AACL ($r = -.01$). In the group of 51 pregnant women, the Month AACL average did not correlate significantly with age ($r = -.12$) or with education as measured by years of school ($r = -.07$).

NORMS

Although the samples used are not large enough to warrant establishing standard scores, the reader may be interested in comparing their samples with the ones used in

TABLE 4
CORRELATIONS BETWEEN AACL ANXIETY
SCORES AND MAS

Group	N	Time Set		
		Today	General	Month
College I	50	.29*	.18	
College II 1st testing	32 ^a	.32	.58**	
M nonexam days		.52**		
M exam days		.28 ^b		
Pregnant: first testing	51			.65**
Pregnant: Month 6-9	10			
Month 6				.72*
Month 7				.85**
Month 8				.77**
Month 9				.38

^a Two Ss did not take the MAS.

^b Exam I $r = .40^*$, Exam II $r = .09$, Exam III $r = .19$.

* Significant at or below .05.

** Significant at or below .01.

this study. The means and standard deviations for the combined college group of 84 Ss were: Today, $M = 7.38$, $SD = 3.98$; General, $M = 5.15$, $SD = 3.06$. For the group of 51 pregnant women given the Month version of the AACL for the first time, $M = 7.14$, $SD = 3.74$. In using these normative figures one should consider the characteristics of the samples as described in the article.

SUMMARY AND CONCLUSIONS

A list of affectively toned adjectives was used in the development of an anxiety scoring key. Adjectives which differentiated High and Low Anxiety groups (selected by psychiatric interview), and showed significant changes in checking frequency during a hypnotically suggested anxiety state, were used in the key. Two forms of the test with different time sets were used in a reliability study. Both the "General" and "Today" forms had adequate internal reliability on a single occasion, but only the General form demonstrated marked retest reliability. These results were expected because of the differences in the time set. The Today version of the AACL was given repeatedly to a second group of college students at the beginning of a class period. Anxiety scores on examination days were compared with scores on nonexamination days. The AACL score was found to rise significantly on examination days. On two of the three examinations, the students who made good grades on the examinations showed a significantly smaller rise in anxiety scores than the students who made poor grades on the examinations. The General AACL correlated significantly with the MAS in one of the two college samples. The mean of 10 nonexamination day AACLs correlated significantly with the MAS. The mean of the 3 examination day AACLs did not correlate significantly with the MAS, although the score on the first examination day did correlate significantly with the MAS. AACL and MAS were highly correlated in a group of pregnant women.

The General version of the AACL is suggested as a quick measure of general anxiety level. The Today version of the test is suggested as a technique in studies where repeated assessments of anxiety must be made within a relatively limited time interval, i.e., weeks or months. The time set of the AACL can be changed by a simple adjustment of the instructions. In the study of pregnant women a "month" set was used. In an experiment taking place over an interval of a few hours a "now" or "this minute" set could be used. It is assumed that the anxiety scoring key will be valid regardless of the time set. The studies reported give some evidence to support this assumption although only future work can establish the fact. Other experimenters may desire to develop new keys for this test. "Depression" and "Hostility" keys are two obvious possibilities.

REFERENCES

- BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327-333.
- GOUGH, H. G. *Reference handbook for the Gough Adjective Check-List*. Berkeley: Univer. California Inst. Pers. Assess. Res., 1955.
- LEVITT, E. E., DEN BREEIJEN, A., & PERSKY, H. The induction of clinical anxiety by means of a standardized hypnotic technique. *J. Amer. Soc. Clin. Hypnosis*, in press.
- NOWLIS, V. The development and modification of motivational systems in personality. In, *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1953. Pp. 114-138.
- PERSKY, H., MAROC, J., CONRAD, E., & DEN BREEIJEN, A. Blood corticotropin and adrenal weight-maintenance factor levels of anxious patients and normal subjects. *Psychosom. Med.*, 1959, 21, 379-386.
- SCHAEFER, E. S., & BELL, R. C. Development of a parental attitude research instrument. *Child Develpm.*, 1958, 29, 339-361.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- ZUCKERMAN, M., RIBBACK, BEATRICE B., MONASHKIN, I., & NORTON, J. A. Normative data and factor analysis on the parental attitude research instrument. *J. consult. Psychol.*, 1958, 22, 165-171.

(Received November 19, 1959)

AN EXTENSION OF THE CONSTRUCT VALIDITY OF THE EGO STRENGTH SCALE

BENJAMIN KLEINMUNTZ

Carnegie Institute of Technology

The present study was designed to add strands to the nomological network (Cronbach & Meehl, 1955) which comprise the Ego Strength (*Es*) scale (Barron, 1953). It was reasoned that if the *Es* scale can differentiate between psychiatric and nonpsychiatric adults (Gottesman, 1959), and if it really measures ego strength, then adjusted college students should get significantly higher scores on this scale than do maladjusted college students. Furthermore, if one of the things which may be said about the *Es* scale is that it reflects a defensive test taking attitude, then it would seem that our sample of adjusted students' defenses would be higher than those of the maladjusted students.

A random sample of 50 college students (27 males and 23 females) was drawn from a group of 300 sophomore and junior Teachers College candidates at the University of Nebraska. MMPIs had been administered to these students as a routine screening device, and it was generally known that the test results had some bearing on their futures in the Teachers College. The criterion of adjustment was met if these students indicated on the Cornell Medical Index that they had never been under psychiatric treatment and if they had never contacted the Mental Health Division at the University of Nebraska. The

maladjusted group of students consisted of 33 individuals (24 males and 9 females) who had voluntarily sought psychiatric consultation at the Mental Health Division and who had remained in treatment for five or more interviews. It is possible of course, with such criteria of adjustment and maladjustment, that a number of really well adjusted individuals might be included in the maladjusted group and that conversely a number of maladjusted individuals might be included in the adjusted group. If there is an error here, however, the error is in the direction of making it more difficult for the *Es* scale to discriminate between the two groups.

The MMPI records of the two groups were rescored for *Es* and *K* and it was found that both scales do tend to broadly discriminate between adjusted and maladjusted college students. These results appear in Table 1.

The mean differences between the two groups on each of the two scales were significant beyond the .001 level of confidence. The mean *Es* scale score for the adjusted students was much higher than that of the maladjusted students; and the mean *K* scale score of the adjusted students was also much higher than that of the maladjusted. Correlations of .32 (high *Es* with high *K*) and .48 (low *Es* with low *K*) between *Es* and *K* for

TABLE 1
Es AND *K* SCORES FOR ADJUSTED AND MALADJUSTED COLLEGE STUDENTS

Group	<i>N</i>	<i>Es</i>	<i>SD</i>	<i>CR</i>	<i>K</i>	<i>SD</i>	<i>CR</i>
Adjusted	50	50.10	4.76	4.82*	19.12	3.96	3.35*
Maladjusted	33	45.21	3.35		15.33	5.53	

* Significant beyond the .001 level of confidence.

PSYCHOLOGICAL TEST REVIEWS

GARDNER, E. F., & THOMPSON, G. G. *Syracuse Scales of Social Relations*. Elementary Level, Grades 5 and 6, Junior and Senior high school level. Test booklets, pkg. of 35 (\$4.90) includes scoring guide, report folder, tally sheets, class record, and manual, 24 pp. Yonkers-on-Hudson, N. Y.: World Book Company, 1959.

This sociometric device makes use of two hypothetical situations as a basis for ratings by each student of his classmates. Approximately parallel forms are developed for use with elementary, junior, and senior high school pupils. One of the situations, at all three levels, involves rating others' ability to offer support, comfort, and sympathy, and is intended to reflect need for succorance; the other is specific to the level. At the elementary level achievement-recognition is tapped; at Junior high, deference, at Senior high, playmirth. The student's frame of reference for his ratings is established in a forced distribution, using individuals selected from all the persons he knows as reference points. Since every pupil is evaluated by every other one, information becomes available on: (a) how each pupil views his classmates as being able to satisfy two of his important psychological needs; (b) how each pupil is evaluated by his classmates as being able to satisfy their needs. Large samples, unspecified with regard to such relevant factors as intelligence and socioeconomic status, provide a not fully satisfactory normative basis for interpreting the average ratings given and received for each need.

An instrument like this one departs from the pattern of the simpler pencil and paper tests. Evaluation of its usefulness becomes an extremely complex process. The authors rely on five years of research, mostly with college students. Test-retest measures of stability over a one-to-two-week period suggest that these measures have only moderate stability (from .62 to .94). The manual is deficient in not warning potential users that such factors as the passage of time or the period in the school year might have considerable effect on a given score. This uncertainty would tend to impair its usefulness for routine diagnostic use. It will not be possible to involve a whole class in the time consuming procedures required every time the need to understand and help a particular pupil is discerned. Its validity is, of course, not susceptible to simple summary. Brief reference is made to studies whose results are suggestive, and the reader is referred to the author's 1956 book and

to a number of unpublished doctoral dissertations. There is lacking any comprehensive or coherent framework for interpreting the results. True, it might be argued that providing such a framework requires more space than a manual can provide. Then the potential user ought to be warned that such a gap exists and referred to the proper sources.

This reviewer is more impressed with possible usefulness of these scales for research than for everyday clinical or educational work.—E. S. B.

DINKEL, R. E. *Survey Test of Algebraic Aptitude*. 8th Grade. Timed, 40 minutes. Test booklets, IBM answer sheets, IBM electronic score punch cards, scoring key; manual, 18 pp. Los Angeles, Calif.: California Test Bureau, 1959.

This 60-item test of basic arithmetical and simple mathematical operations is designed to be used as a basis for guidance and placement for ninth grade mathematics. It seems to have undergone adequate standardization and some degree of preliminary item analysis. Its reported predictive power against achievement tests scores or grades earned in ninth grade mathematics is satisfactory. The estimated reliability is probably inflated through the inappropriate application of Kuder-Richardson Formula 21 to a speeded test. Considerable correlation with similar tests (Lee, Iowa, and Orleans tests) are reported, but no comparisons with their respective predictive validities are offered. This test may prove as dependable as the others.—E. S. B.

JASTAK, J. F. *Jastak Test*. (Short form) Junior high school, 1 form. 36 minutes. Test booklet (\$3.85 per package), set scoring key (\$.75), manual, 61 pp. (\$1.50). Philadelphia: Educational Test Bureau, 1959.

This briefer form of the 1958 edition (see *J. consult. Psychol.*, 1959, 23, 283) is reported to correlate .97 with its parent. This correlation and the extremely high reliability coefficients reported are fallaciously high estimates because they were computed on the same sample that provided the data for item analysis on the basis of which the items to be included in this form were selected. The manual is essentially unchanged. I must repeat my conclusion regarding the long form: without further validation data, the *Jastak Test* cannot be accepted as a way of obtaining inferences other than general level of ability.—E. S. B.

CLIENT ACCEPTANCE OF RESPONSIBILITY AND DIFFICULTY OF THERAPY

PEARL SCHROEDER

Student Counseling Service, University of Illinois

This study was designed to test the proposition that there is a meaningful relationship between the degree of responsibility demonstrated by clients and the difficulty they pose for therapeutic work. Difficulty is here limited to its more extreme forms and refers either to the instance in which, although a person requests help, he does not really "get into" therapy, or to the instance in which many more hours than usual must be spent to achieve the therapeutic goal.

Clinical observation suggested that one common dimension of difference among clients of varying complaints and symptomatology is the way in which they view themselves as contributors to their experience. On the one hand there is the client who presents himself as the helpless recipient of the impact of other people and of chance events; on the other, the client who claims himself as the cause of his difficulties. Such difference appears not only to operate in initial presentation but also to permeate general speech content and to be reflected in the client's role expectations of himself and the clinician in the therapeutic work. Abstracted from the varying contexts in which it was noted, this difference can be identified as a variation in the extent to which a client views his own attitudes and behaviors as being consequential, that is, as playing a part in the course of events. Stated dictionary-wise, it is a difference in the extent to which a client holds himself answerable as an agent for these events or is "responsible."

Theoretical discussions of personality and neurosis associate the rejection of responsibility with repression and the development of psychological distress, and the acceptance of responsibility with normal development

and psychological well-being. To illustrate: Thompson (1958), speaking of the development of the self, notes that "The 'good me' tends to deny responsibility for the 'bad me's' activities" (p. 9) and that trouble ensues; and Horney (1950), describing the neurotic, states that "he always tends to deny laws operating within himself, refuses to see the inevitability of cause and effect operating in psychic matters . . ." (p. 36). Rogers (1951) speaks of the ill effects of a failure to "own" experiences and then links responsibility to the normal development which moves "in the direction of greater independence or self-responsibility" (p. 488). Fingarette (1955), in an examination of Freud's position anent moral guilt and responsibility, after linking repression and the rejection of responsibility, concludes that "accepting responsibility is a necessary condition of maturity" (p. 32).

Fingarette's discussion is specific to "moral" responsibility. In general use, "responsibility" does often carry the additional meaning of some form of moral judgment such as guilt or blame. This additional meaning may be a necessary consequent in terms of personal and social values. However, the claiming of blame or the making of the moral judgment is separable as a process from the claiming of responsibility defined above as viewing one's own behavior as consequential. Fingarette points out that guilt which seems disproportionate to the assigned reasons is but guilt dissociated from behaviors (wishes) which are not avowed. Thus there is a claiming of the moral judgment but a disavowal of the cause for the effect. In the statement, "Registration happens to me tomorrow," no moral judgment is involved, yet the speaker implies he simply experiences what some outside force does to

him and rejects viewing registration as a requirement whose fulfillment depends upon his activities.

Insofar as rejection of responsibility is associated with rejection of reality (Horney, 1950; Thompson, 1958) and assumption of responsibility with development and maturity (Fingarette, 1955; Rogers, 1951), a simple linear relationship between client acceptance of responsibility and difficulty of therapy would appear to be the most logical one. However, clinical observation suggested that the client who attributes his difficulties almost totally to himself, later in therapy has as much difficulty in accepting and dealing with reality relationships as the client who attributes his difficulties almost totally to outside forces. Logically, too, it can be argued that the person who rejects the effects of outside influences and claims total responsibility for the events of his life is in this respect divorced from reality and will therefore be difficult to work with in therapy. These last considerations led to the prediction of a curvilinear rather than a linear relationship and the final form of the hypothesis to be tested was: clients showing extremes of responsibility pose more difficulty for therapy than those showing moderate responsibility.

METHOD

Measures

Two measures were required: one of responsibility, the other of difficulty. In both instances a more objective method than global clinical judgment was desired. Since the observations leading to the hypothesis were of what the client said, a content analysis of verbal production was chosen for the measure of responsibility. Personal characteristics can be demonstrated by very brief verbal material; however, identification of such characteristics for later recognition requires material of some length. Therefore a fairly lengthy protocol was necessary.

A three-years' collection of data for clients who had participated in an extensive research project in therapeutic counseling sponsored by the Student Counseling Service and the Department of Psychology of the University of Illinois had just been completed (Hunt, Ewing, LaForge, & Gilbert, 1959). The data included protocols of therapy interviews, Rorschachs, TATs, and Gilbert Self-Interview Test.

Because the situation in the Gilbert Self-Interview Test is simpler than that of the interview with its dyadic interactions, and because the subject (S) of necessity must focus more directly on himself and his

life events than in the Rorschach or TAT, the Gilbert protocols were chosen for the content analysis.

The Gilbert Self-Interview Test (Ewing, 1957, pp. 19-20) is administered by written instructions. The S is alone in a room and speaks his responses into a microphone which records them on tape. These responses are then transcribed and the tapes filed for listening. The test asks S to spend about three minutes in describing his "real attitudes and feelings both good and bad" and his "problems, if any" about each of the following 10 topics: educational and vocational goals; family; how he gets along with people of both sexes; relationships with the opposite sex; financial situation; ethical, moral, and religious views; abilities; unfavorable and undesirable aspects of his personality; favorable and desirable aspects of his personality; any important aspect of his life not covered. The resultant protocol is generally of some length.

Instructions for the analysis of these protocols do not presuppose any psychological training, nor do they require any clinical interpretation. They are but briefly summarized here.¹ The unit of analysis is the statement, a group of words expressing a separate idea. All statements of the following kinds are identified and marked as responsibility statements: (a) reasons given for an event or condition, (b) statements indicating the influence or effect of an agent upon a recipient, (c) statements of cause and consequence, (d) identification of means used for accomplishing ends, and (e) statements of need or compulsion arising from without or from within. Those responsibility statements in which S presents himself as the active agent are then scored plus; those in which he presents himself as the recipient of an agent's activity, minus; and those not classifiable as either of the above, check.

Reliabilities of scoring were established for two judges on 10 cases. Since the scores of both judges are measures of the same thing and therefore means and standards deviations should be equivalent, the intraclass correlation measure described by Haggard (1958) was used to estimate reliability. R values for the raw scores were: plus scores, .915; minus scores, .684; check scores, .778; sum of all responsibility statements, .936. The probabilities associated with these correlation values are in every instance less than .01.

To measure responsibility as the extent to which S views his behavior as consequential, it was necessary to take into account those instances in which he disclaims his own agency and attributes the source of events to something outside himself. Therefore the

¹ A complete copy of the instructions for the use of the termination form and for assessing difficulty has been deposited with the American Documentation Institute. Order Document No. 6406 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

responsibility score had to represent both plus and minus scores. Three such compound scores were possible: two ratio scores and one difference score. The two ratios were: number of plus responsibility statements/sum of all responsibility statements, and number of plus responsibility statements/sum of plus and minus responsibility statements. The difference score was the number of plus statements less the number of minus statements corrected for sign by the addition of a constant. The intraclass correlation for the two judges was computed for each of these scores with the following results: ratio score of plus statements/sum of all responsibility statements, $R = .671$, $p < .01$; ratio score of plus statements/sum of plus and minus statements, $R = .677$, $p < .01$; difference score (number of plus statements less the number of minus statements), $R = .591$, $p < .05$.

Arbitrarily, the ratio of sum of plus/sum of all plus and minus statements was chosen as the measure for the test of the hypothesis. Extreme and moderate responsibility was determined by quartile divisions of the distribution of these scores. Not as a part of the test but as a part of exploratory work, it was planned to subject the other two compound scores to the same statistical analysis as the plus/plus and minus ratio score.

A two-point scale of difficulty was established. Difficulty has already been defined with reference to those instances in which a person either does not really "get into" therapy or he requires a "long time." The empirical measure of difficulty was based upon the number of interviews and upon data supplied by the termination form (see Footnote 1) which, filled out by the therapist at the end of therapy, describes the conditions of termination.

The category "difficult" included all cases with more than 30 interviews (roughly two semesters' interviews on a once-a-week basis) and all those terminated after fewer than 10 interviews with termination-form data which indicated that, in the therapist's opinion, termination was dictated by resistance or lack of motivation rather than completion of therapeutic work. "Nondifficult" included all those cases not specified above.

Design

Using the two measures described above, the empirical prediction for testing the hypothesis becomes: a significantly greater proportion of those cases with extreme responsibility scores (sum plus statements/sum plus and minus statements) are associated with difficult therapy than with nondifficult.

The statistical test was a two-by-two chi square.

Subjects

The 48 Ss who constituted the sample for this study were all clients of the Student Counseling Service who had participated in the therapy research project conducted by the Counseling Service and the Department of Psychology of the Univer-

sity of Illinois.² This service is available to students, both undergraduate and graduate, and, to a limited degree, to faculty and to nonacademic personnel. The service has no hospital facilities and the client population is of necessity limited, therapy-wise, to those who can manage on an outpatient basis.

The sample was selected on a number of bases. Since termination data were necessary for the categorization of difficulty, only those clients who applied for service and had at least one therapy interview could be used. Ten such cases, which were also used for the establishment of reliability, were pulled at random from the file of typed transcripts. Eleven were chosen as part of an initial effort to limit the number of different therapists involved (11 different therapists contributed to the termination data). Eleven cases were selected to maximize the overlap of Ss between this study and two others and so to provide interstudy information. Sixteen were selected because posttherapy Gilbert Self-Interview Test protocols were available for them and a pre-post study had been tentatively planned.

Although the sample was neither systematically selected on a single set of variables nor completely randomized, the method of selection of cases does not appear to foster any bias except an increase in the number of Ss with posttherapy test data over what it might have been with complete randomization. (Thirty-one of the sample repeated the Gilbert test after termination of therapy.)

The age range of the sample was 17 to 32 with a median of 24. Number of interviews per client ranged from 2 to 118 with a median of 21. The 48 Ss divided evenly between the difficult and nondifficult categories, 17 men and 7 women falling into each category.

RESULTS

A two-by-two chi square test of the distribution of Ss with extreme and with moderate responsibility scores (plus statements/sum plus and minus statements) between the difficult and nondifficult categories showed no significant relationships ($\chi^2 < 1.00$). The same test applied to the other two compound scores also showed no significant curvilinear relationships between difficulty and responsibility.

A check of the rejected linear relationship, in which difficulty of therapy would have been associated with low responsibility, showed

² The program of research which supplied the raw data upon which this study is based was initiated under a grant to J. McV. Hunt, T. N. Ewing, and W. M. Gilbert from the Carnegie Corporation. It was also supported by the Research Board of the University of Illinois, and the major share of support has come from grants from the National Institute of Mental Health (M-1041 and M-1388) and a contract with the Office of Naval Research (Nonr-1834).

TABLE 1

CHI SQUARE TESTS OF ASSOCIATION: RESPONSIBILITY, DIFFICULTY OF THERAPY, AND MOVEMENT

	D	M
R	$\chi^2 = 6.761^{**}$ High R with D	$\chi^2 = 4.090^*$ High R with High M
D		$\chi^2 = 5.333^*$ D with High M

Note.—All chi square values based on 2×2 tables.

R: Responsibility as measured by number of plus statements less the number of minus statements. Median splits used to determine high and low.

D: Difficulty of therapy as determined by length and conditions of termination. Cases divided into difficult and nondifficult.

M: Movement judgments made by therapists at termination of therapy. Median splits used to determine high and low.

* $p < .05$.

** $p < .01$.

that statistical significance depended upon the particular compound score used but that the direction of relationship was the opposite of what would have been predicted from the literature. The results of these analyses, using two-by-two chi squares with median splits for the responsibility scores, were: ratio score of plus statements/sum of all responsibility statements, $\chi^2 < 1.00$; ratio score of plus statements/sum of plus and minus statements, $\chi^2 = 3.00$, $p < .10$; difference score (number of plus statements less the number of minus statements), $\chi^2 = 6.761$, $p < .01$. The last two show difficulty of therapy associated with high, not low, responsibility scores.

A different criterion variable, movement judgment, had been used in the large therapy research project carried on by the Student Counseling Service (Hunt et al., 1959). This judgment, based on the Hunt-Kogan Movement Scale (1950), was made by the therapist at the conclusion of therapy and was available for every S of the present study.

To explore the significance of the observed relationship, the relationships of responsibility (difference score) and difficulty to this second criterion, movement, were analyzed. The results are summarized in Table 1. They show that high responsibility is associated with difficult therapy, that high responsibility is

associated with high movement, and that high movement is associated with difficult therapy.

Seventeen cases showed concomitance of high responsibility and difficult therapy; of these 17, 11 also showed high movement. Ten of these 11 were classified as difficult because of excessive length of therapy. (Of the 24 difficult cases, 19 were so classified because of excessive length, and only 5 because of early termination with rejection of therapy.) Sixteen cases showed concomitance of low responsibility and nondifficult therapy. Of these 16, 13 also showed low movement.

The relationships of Table 1 are thus largely determined by two kinds of cases: those with high responsibility scores who pose the difficulty of excessively long therapy but show high movement; and those with low responsibility scores who neither flee therapy early nor take up much time but do not show much movement.

DISCUSSION

Although the data do not support the hypothesized curvilinear relationship nor yet the obvious inference from the literature of a linear relationship with low responsibility associated with difficult therapy, they do demonstrate an association of these two variables which cannot be ignored.

Exploration of the meaning of these results by means of an analysis of the relationships of these two variables to a second criterion, movement, showed that high responsibility scores, difficult therapy, and high movement were all positively associated with each other. As pointed out in the presentation of results, the high part of this three-way relationship depends largely upon those clients who pose the difficulty of excessively long therapy; the low upon those who neither flee nor take a long time, but who do not show much change. Therefore it is suggested: that clients with high responsibility scores, if they do not flee, ask more of therapy, are willing to invest more in it, and, as a result, both take more time and show more change; and that clients with low responsibility scores have a different, perhaps a less self-focused goal, limit what they will invest, and therefore take less therapy time and also show less change.

One implication of the present results for replication and further investigation concerns the measure of responsibility. The fact that the two scores which are most narrow in scope are also the most discriminative with respect to the criterion suggests it would be worthwhile to try to simplify and sharpen the instructions for content analysis. Another implication concerns the whole problem of predicting therapy but can be best articulated here by focusing upon cases with high responsibility scores: what differentiates those cases with high responsibility scores who persist and gain from those who do not? It is suggested that motivation is such a variable and should be studied in conjunction with responsibility for predicting the course of therapy.

SUMMARY

It was hypothesized that clients with both very high and very low acceptance of responsibility would present difficulty for therapy when such difficulty was defined as either never really "getting into" therapy or taking a greater than usual number of interview hours. To test this hypothesis, a method of measuring responsibility by means of content analysis of verbal protocols was developed.

The data did not support the hypothesized curvilinear relationship. They showed instead an association of high responsibility

with difficult therapy and low with nondifficult. They also demonstrated a relationship of high movement with high responsibility, and high movement with difficult therapy.

It was suggested for further work that the measure of responsibility be sharpened, and that motivation be investigated in conjunction with responsibility for predicting therapy outcomes.

REFERENCES

- EWING, T. N. Therapeutic counseling research project: Research Report No. 1. U. Illinois: Student Counseling Service and Department of Psychology, 1957. (Mimeo)
- FINGARETTE, H. Psychoanalytic perspectives on moral guilt and responsibility: A re-evaluation. *Phil. phenomenol. Res.*, 1955, 16(1), 18-36.
- HAGGARD, E. A. *Intraclass correlation and the analysis of variance*. New York: Dryden, 1958.
- HORNEY, KAREN. *Neurosis and human growth*. New York: Norton, 1950.
- HUNT, J. McV., EWING, T. N., LAFORGE, R., & GILBERT, W. M. An integrated approach to research on therapeutic counseling with samples of results. *J. counsel. Psychol.*, 1959, 6, 46-54.
- HUNT, J. McV., & KOGAN, L. S. Assessing the results of social case work: A manual on judging movement. New York: Family Service Association of America, 1950.
- ROGERS, C. R. *Client-centered therapy*. New York: Houghton Mifflin, 1951.
- THOMPSON, CLARA. Concepts of the self in interpersonal theory. *Amer. J. Psychother.*, 1958, 12, 5-17.

(Received November 20, 1959)

SOCIAL VALUE-NEED PATTERNS IN MENTAL RETARDATES¹

THOMAS E. BOLDUC²

Laurelton State Village, Pennsylvania

One of the important factors that characterizes the behavior of socially adjusted individuals is the awareness of and ability to conform to social values. As individuals mature, society expects them to acquire an increasingly conscious and rational code of behavior. Certain minimal standards of socially desirable behavior must be acquired to achieve a satisfactory social adjustment. In a sense it is the ability to cope with everyday life experiences without opposing society's values that constitutes an acceptable social adjustment.

The increasing complexity of the process of living in modern cultures augments the handicap for the person of limited intelligence. His ability, for example, to acquire norms of behavior for a wide variety of situations, to apprehend the character of his experiences, and to form comparisons between them is more limited and less flexible than that of a person of normal intelligence. Since he must form ideas of situations he has never actually experienced and conceive of events and how to meet them before they ever occur, it is little wonder that the mental retardate becomes a potential delinquent. The inability to acquire and apply socially desirable behavior may be one of the most important reasons for this failure in community adjustment.

The purpose of this investigation was to isolate and identify common patterns of solutions to social value-need conflict situations in mental retardates. To achieve this goal an in-

strument was constructed to measure the responses to social value-need conflicts in situations similar to those which might occur in everyday experiences of mental retardates. Through an inverse factor analysis of correlations between persons on this measure, patterns representing groups who favor particular types of solutions to these conflicts were sought.

THE RESEARCH DESIGN

Factor analysis was considered to be the most effective method of identifying the multi-dimensional variables associated with such general concepts of values and needs. Although the study was exploratory, the explorations were based on definite principles with a specific structure designed to define the limits of the area of investigation.

A review of the literature failed to uncover a standardized scale that would adequately measure the variables of interest in this study. Instruments (Allport, Vernon, & Lindzey, 1951; Anderson & Anderson, 1954; Bills, Vance, & McLean, 1951; Grace & Grace, 1952; Johnson & Stanley, 1955; McDonald, 1956; Morris, 1956; Woodruff, 1942) that have been developed to measure value structure would have required modifications and extensive editing to adapt them for use with a population of mental retardates. It was therefore necessary to design a measure based on a classification structure which would reflect individual differences in the mental retardate's responses to social value-need conflict situations.

Selection of the Variables

The first problem was to select appropriate variables for the two major dimensions of

¹ The research was supported by a USPHS research grant (M2121) from the National Institute of Mental Health, United States Public Health Service. Appreciation is expressed to Alfred Butler, project director, for his editorial suggestions and valuable critique which were very helpful in the preparation of this paper.

² Now at Psychiatric Services, Wales, Wisconsin.

value and need, and then to devise a classification system that would facilitate the construction of the items and provide structure for the scale.

In keeping with the goal of identifying patterns of social value-social need conflicts in familiar situations, it was necessary that both the value and the need dimensions include social rather than personal areas. Even with this limitation imposed on the structure a wide variety of possible classifications existed. It must be recognized that the concepts of value and need are not independent psychological entities but complex multidimensional variables that overlap other personality variables. Therefore, the specific values and needs selected for this study are "social" only in terms of emphasis. It is conceivable, for example, that an individual could be honest or loyal to his self-system, but it is more common to consider him honest or loyal to others. Similarly, the needs selected are primarily social rather than personal in that their gratification depends primarily on interactions or relationships with others.

The classification system of values employed in Havighurst and Taba's (1949) study of character and personality met the above criteria and provided a comprehensive coverage of the value dimension required for the present investigation. The categories of values used were *honesty, loyalty, moral courage, and responsibility*. Other values could have been employed, but after reviewing the literature in this area it was apparent that many of the value descriptions were related to or were modifications of the values used in the Havighurst study.

Honesty as employed in this study refers to the use of property and truthfulness. Being truthful, especially with authorities such as parents, teachers, and employers, and having respect for the property of others are examples of honesty. Loyalty involves being faithful to school functions and staff, to friends who are in trouble, and to one's family. Moral courage involves the protection of one's own rights and those of others. The defense of the self and others, including minority groups, against harmful gossip, the expression of personal opinions and beliefs particularly when resistance is encountered, and

"preachy behavior" are examples of moral courage. Responsibility involves the honoring of home, school, and employment duties. The completion of assigned tasks, family responsibilities, and the protection of people or property under one's care reflect responsibility.

In a similar manner, needs were chosen that met the criteria and yet were comprehensive enough to cover a wide range of familiar, everyday experiences. Five needs, based on Murray's (1938) theoretical framework, were employed in the present study. They are *nurturance, succorance, dominance, affiliation, and deference*. All of these imply that a relationship with another person facilitates gratification. Considerable research in many areas of study has been reported, utilizing these needs. Edwards (1953), for example, includes these five needs in his study of personal preferences.

Nurturance is defined as the need to give affection, sympathy, help, or pity to another person. Succorance is the need to be nursed, supported, protected, loved, and consoled by another person. Dominance is the need to control, influence, or direct the behavior of others either by suggestion, persuasion, or command. Affiliation involves the need to form or maintain a relationship with another person. Deference involves the need to yield, submit, or refuse to compete with superiors because of physical size, status, or age.

The Research Instrument

A pool of 141 items was constructed according to the criteria described above. Twenty pairings were possible so that there were items describing a conflict between each combination of the four values and the five needs. Each item consisted of a brief description of a familiar social conflict situation and a forced-choice decision between two alternate responses—one depicting a value and the other a need. The forced-choice technique reduces the effect of such biasing variables as response set. The items were typed on 3 × 5 cards with the order of alternates randomized to reduce the possibility of a position set occurring. An attempt was made to include a comprehensive range of social experience familiar to mental retardates. The task of the subject (S) was to select one of the alternate solutions as the response to the conflict situation. In order to minimize the anxiety associated with the administration procedures as well as to gain insight into the social acceptability status of the items it was decided that the Ss adopt a "should" rather than a

"would" set. It was hoped that this would reduce the defensive test behavior associated with making self-statements. Examples of the items are presented below.

#29. Leona was in a gang. One day the gang broke some windows at school. Leona was caught and the other girls got away. The teacher asked Leona who the other girls were. What should Leona do? (b) Tell the teacher because teachers are the boss. (a) Not tell the teacher because she does not want to get the other girls into trouble. (Structure: Loyalty vs. Deference)

#56. Sue had a job checking spelling papers after school. One day a boy came by and asked Sue to go to the movies with him. What should Sue do? (a) Keep on with her job. (b) Go with the boy because Sue needed someone to like her. (Structure: Responsibility vs. Succorance)

#60. Blanche was a lonely girl. Blanche's report card had some bad marks on it and she was afraid to take it home. What should Blanche do? (b) Change the marks and make them higher so that her father would be pleased with her. (a) Take it home and show it to her father. (Structure: Honesty vs. Succorance)

#86. Joan was one of the best players on the club's kickball team. One day the girls from another team told Joan that if she would play for them they would make her captain. What should Joan do? (b) Go with the other girls and be the captain. (a) Stay with her own team because they needed her. (Structure: Loyalty vs. Dominance)

After the items were constructed and refined they were submitted independently to four psychologists who were instructed to identify each item according to the criterion definitions of the classification structure. A total of 21 items failed to meet the 75% interjudge agreement criterion and were eliminated from the scale. It should be mentioned briefly that this method measures only the degree to which the items are descriptive of the criteria definitions. The future use, therefore, of the terms "value" and "need" will refer only to the judged face validity as empirically determined.

Following this refinement, the remaining 120 items were administered on two occasions, seven days apart, to a random sample of 20 institutionalized females, with an age range of 14 to 18 and an IQ range of 57 to 78, none of whom had been institutionalized more than two years.

The test-retest reliability coefficients (tetrachoric) for each S were transformed to z scores and averaged, resulting in a mean test reliability of .87. Similarly the reliability coefficients of each of the 120 items were averaged, resulting in a mean item reliability of .93. Eleven items were eliminated due to low or negative reliability coefficients.

A criterion of 10% or greater was adopted to determine the variability measure for each item. Fourteen items failed to discriminate 10% or more of the sample, and were eliminated. Each of the remaining 95 items which comprise the final scale³ met the reliability and variability requirements.

As might be expected, the item distribution was skewed slightly toward the value dimension in that the Ss tended to select the value alternate more frequently than its need counterpart. For this reason some of the biased value alternatives were de-emphasized, that is, made less socially desirable. The above panel of judges agreed that this moderate rewording of some of the items should not significantly affect the reliability and should increase the variability of the item.

The final distribution of the items according to the classification structure seemed adequately balanced. No single category had less than three replications or more than seven, with a mean of five. The Ss were able to comprehend the items and meaningfully select alternates. It was felt that the Ss were able to maintain the "should" set and that the defensiveness associated with examinations was decreased.

The Subjects

The Ss in this study were 50 mental defectives, consisting of two groups of 25 each, between the ages of 14 and 18 and with an IQ range of 55 to 80. One group was institutionalized and the other was not institutionalized. All the Ss were without gross organic dysfunction and only white females were used, to eliminate race and sex differences in response.

The institutional sample of 25 retardates was drawn randomly from the total population of Ss meeting the age and intelligence criteria at the Laurelton State Village. This is a state training school for female mental retardates of child-bearing age, the majority of whom are court committed because of a failure to achieve a satisfactory social adjustment in the community.

The noninstitutional sample was drawn randomly from the known retardates of the above-specified age and IQ range in three public school districts. The population from which the sample was drawn approximated the institutional population in rural-urban distribution. None of the Ss selected had been involved in court action for delinquent behavior.

Although the chief objective of the study was not to measure response differences between groups of Ss, it was decided that if two random groups of Ss could be obtained which were similar in all respects except institutionalization, it would be feasible to make inferences about differences that might point the way for future research. If only institutionalized Ss, for example, were used the findings of the study would be restricted to institutionalized retardates. It was decided that greater variability in social value-need behavior would be obtained if noninstitutionalized Ss were included in the sample. For these reasons random samples were selected and tests of significance of differences were computed for those variables that might confound the data. No significant differences were found in age, intelligence, or reading ability.

³ This scale has been deposited with the American Documentation Institute—see Footnote 5.

Testing Procedure

The scale was administered individually to each of the 50 Ss. Each item was presented verbally to the S who was requested to identify the better one of the two alternative solutions to the conflict situation depicted by the item. None of the Ss refused to take the test and all Ss tested were contained within the originally selected sample.

Statistical Analysis

The major hypothesis of this study was that psychologically meaningful dimensions would be revealed if a matrix of intercorrelations of Ss' responses to pairs of alternate solutions was factor analyzed and the arbitrary orthogonal axes rotated to simple structure. The principal axis method of factor extraction (Hotelling, 1933) and the orthogonal rotation method were chosen as the methods of analysis. This method provides a more rigorous means of extracting factors from a correlation matrix and it has the advantage of placing more of the variance in the first factors. The highest correlation coefficient in the column was used as the initial estimate of communality. Subsequent estimates were based on the previously obtained communality.

Each S's responses were correlated with those of every other S, resulting in a 50×50 matrix of intercorrelations. Since the data were dichotomous and the underlying dimensions were assumed to be continuous, the tetrachoric coefficient was considered most appropriate. An approximation of the tetrachoric was made possible by programing PENNSTAC,⁴ utilizing the first five terms of the tetrachoric equation. Extraction of factors was discontinued after six factors. Residual correlations ranged from $-.2938$ to $.2713$, with a mean of $.0006$ and a standard deviation of $.0877$. Rotation of the arbitrary orthogonal reference axes was performed by the varimax solution (Kaiser, 1956). This solution provides an objective method of approximating simple structure. A comparison of the varimax and the quatrimax rotation methods by Wrigley, Saunders, and Neuhaus (1958) indicated that the former has the advantage of providing a somewhat more stable structure when the number of factors is small and the number of tests (persons) is large, as is the case in this problem. Biserual correlation coefficients were computed to measure the relation of each item to each final factor loading.

RESULTS

The rotated factor matrix of the Ss' responses to the social value-need scale is presented in Table 1.⁵

This table provides the rotated factor loadings on each of the six factors for the 50 Ss. In the first 25 rows are the factor loadings of the noninstitutional sample; the last 25 rows represent the Laurelton sample. The distribution of the total variance accounted for by the six factors is given in the last row of the rotated factor matrix.

The six factors represent six possible patterns of preferences of the 50 Ss obtained through their responses to the forced-choice items of the scale. Therefore, the characteristics of the items which differentiate the Ss on the six factors could serve as a basis for factor interpretation. Accordingly, each of the 95 items of the research instrument was correlated with the factor loadings of each of the six rotated factors. The 570 biserial correlation coefficients obtained were employed to measure the contribution of each item to the rotated factors. The highest item-factor correlations provided the basis for the psychological interpretation of the factors. Items that correlated highly with only one factor were given the greatest weight and provided the identification of the specific nature of the factors. Where items correlated significantly with more than one factor, such items were given less emphasis and were used primarily to define the general nature of the factors.

Since the items were designed so that a forced-choice response between two alternatives was required, it follows that when an S selected the value alternative she was at the same time rejecting the need alternative. Therefore the interpretation of the dimension associated with the preference is inferred not only from the chosen alternative but also from the other alternative in the situation. In some instances the preference for one selection may represent an avoidance response—an avoidance of the other alternative. This

⁵ The social value-need scale, the correlation matrix, the original factor matrix, the transformation matrix, the item-factor correlations, and the table of items defining the factors have been deposited with the American Documentation Institute. Order Document No. 6404 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

⁴ A medium-speed digital computer at Pennsylvania State University. Special thanks are due Donald Laird for his cooperation in developing the program.

correct another person's misbehavior, it is better to feel sorry and accept it. For example, Item 87 indicates that it is better to forgive a friend for stealing than to report her to the authorities. Thus, there is an element of in-group cohesiveness similar, perhaps, to honor among thieves, where there is a sort of mutual protection against authority figures and their rules and regulations.

As might be expected, Ss with the highest loadings on this factor are found principally in the Laurelton sample. This is substantiated by the fact that of the five Ss with loadings of .40 or greater, four are in the institutional sample. The large number of Ss with negative factor loadings in the noninstitutional sample further supports this opinion. This is understandable when one considers that it is the rebellious and negative attitude toward authority in general that has been postulated as a major factor in delinquency. This factor, then, would seem to differentiate individuals who allow their own needs for pleasure and group enjoyment to interfere with the adherence to adult requests, from those who respond diligently and appropriately to such situations. The presence of negativism and rebelliousness toward adults is perhaps the most prominent characteristic of this factor.

DISCUSSION

The chief purpose of the study was to isolate and identify common patterns of response to social value-need conflict situations. This was achieved through a factor analytic procedure and the extracted factors were interpreted.

The factors, however, were extracted from a matrix of intercorrelations between individuals from two combined samples. Thus, the factor loadings of the two samples are not independent, since all the Ss were intercorrelated. The samples were tested, however, for any differences in variables that might confound the data, such as age, intelligence, achievement, and socioeconomic status. The items were originally designed and judged for the absence of inherent differences in the items themselves which might favor one group. If these precautions were adequate, then any differences observed in factor structure between the two groups would be influ-

enced by the presence or absence of the variable of institutionalization.

Although the design of the study does not permit a statistical analysis of group differences in factor loadings, the observed differences, nevertheless, may be useful in a tentative evaluation of the function of the fact of institutionalization.

Factor I, the factor that extracted the greatest amount of variance, and Factor III seem to be weighted by individuals found in the noninstitutional sample. That is, Ss who have yet to come into conflict with authority to the extent that institutionalization would be warranted possess the highest factor loadings on these factors. It would seem then that the interpretation of these factors would be important in the process of considering applicants for commitment.

Factors II and IV do not appear to discriminate between the samples and are probably not directly related to the variable of institutionalization.

Factors V and VI are weighted by individuals in the Laurelton sample. That is, Ss who have been committed to an institution possess the highest factor loadings on these factors. The majority of Laurelton students are committed primarily because they have become a social problem in their community. Therefore, a test heavily weighted with Factors V and VI may be useful in discriminating institutional and noninstitutional samples.

The preceding analysis suggests the feasibility of developing these measures to discriminate subsamples of retardates. For example, the selection of parole students from institutionalized populations might be facilitated. Similarly, a screening device for retardates attending special education classes in the public schools could be developed.

SUMMARY

The purpose of this investigation was to determine the nature of common patterns of solutions to social value-social need conflict situations in mental retardates. To isolate and identify the common patterns of conflict solutions, a scale consisting of 95 items, devised to measure social value-social need conflicts, was administered to 50 white, female retardates between the ages of 14 and 18 with

an IQ range of 55 to 80. Half of these Ss were drawn from an institutionalized population in a state training school for female retardates. The other 25 were drawn randomly from a population of retardates of similar age, intelligence, race, and socioeconomic status attending special education classes in 12 public schools. Each item was constructed so that the S was required to make a forced-choice response, on the basis of what was believed to be the best choice, to a familiar, social conflict situation.

One choice of response was based on a social value and the opposite choice was based on a social need. The social value dimension was arbitrarily divided into honesty, loyalty, moral courage, and responsibility. The social need dimension consisted of nurturance, suc-corance, dominance, affiliation, and deference. Each value was paired with each need, resulting in 20 possible combinations of situations.

The intercorrelations of the Ss on their responses to the items were factor analyzed and orthogonal rotations revealed six definable factors.

Factor I stressed a passive, compliant approach to authority figures with responsibility being the most dominant value orientation. Factor II emphasized loyalty and the protection of others with loyalty being the most dominant value orientation. Factor III was tentatively defined as a preference for submission in situations involving authority figures but with more assertive behavior in relationships with peer groups. Factor IV stressed honesty and moral courage, suggesting an assertive, dominant individual who prefers to participate in the correction of the misbehavior of others. Factor V emphasized the need to be dominant in relationships with peer groups. Values are recognized, but are frequently distorted and employed as weapons with which to gain dominance. Factor VI

stressed a negative, rebellious approach to social situations involving authority figures with an underlying need for affection and pleasurable activities dominating behavior.

Some implications of the factors were discussed. The feasibility of employing this or a similar instrument to differentiate institutional and noninstitutional retardates was explored.

REFERENCES

- ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Study of values*. (Rev. ed.) Boston: Houghton Mifflin, 1951.
- ANDERSON, H. H., & ANDERSON, GLADYS L. Children's perceptions of social conflict situations: A study of adolescent children in Germany. *Amer. J. Orthopsychiat.*, 1954, 24, 246-257.
- BILLS, R. E., VANCE, E. L., & MCLEAN, O. S. An index of adjustments and values. *J. consult. Psychol.*, 1951, 15, 257-261.
- EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychol. Corp., 1953.
- GRACE, GLORIA L., & GRACE, H. A. The relationship between verbal and behavioral measures of value. *J. educ. Measmt.*, 1952, 46, 123-131.
- HAVIGHURST, R. J., & TABA, H. *Adolescent character and personality*. New York: Wiley, 1949.
- HOTELLING, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, 1933, 24, 417-441.
- JOHNSON, O. G., & STANLEY, J. C. Attitudes toward authority of delinquent and nondelinquent boys. *J. abnorm. soc. Psychol.*, 1955, 51, 712-716.
- KAISER, H. F. The varimax method of factor analysis. Unpublished doctoral dissertation, University of California, 1956.
- MCDONALD, F. J. Factors related to moral judgments of children. Unpublished doctoral dissertation, Stanford University, 1956.
- MORRIS, C. *Varieties of human value*. Chicago: Univ. Chicago Press, 1956.
- MURRAY, H. A. *Explorations in personality*. New York: Oxford Univ. Press, 1938.
- WOODRUFF, A. D. Personal values and the direction of behavior. *Sch. Rev.*, 1942, 50, 32-42.
- WRIGLEY, C., SAUNDERS, D. R., & NEUHAUS, J. O. Applications of the quatrimax method of rotation to Thurstone's primary mental abilities study. *Psychometrika*, 1958, 23, 151-170.

(Received October 12, 1959)

SOCIAL DESIRABILITY AND THE FORCED CHOICE METHOD

MARVIN J. FELDMAN AND NORMAN L. CORAH¹

University of Buffalo

When subjects (Ss) respond to personality questionnaires, it cannot be assumed that the test instructions serve as the only or even the major set followed by the Ss. Other self-imposed sets may bias test responses. Meehl and Hathaway (1946) and Cronbach (1946) have both cogently described how test taking attitudes intrude upon personality tests. More recently, Edwards (1957) has stressed the importance of social desirability (SD) stereotypes as a biasing factor. He seems to suggest that this particular response set is so pervasive that it impairs the validity of most standard personality inventories. For Edwards, the forced choice technique affords a possibility of minimizing the effects of SD. His own scale, the Edwards Personal Preference Schedule (EPPS), is constructed accordingly. The forced choice format calls for responses to pairs of statements; the pairs presumably have been equated for SD on the basis of prior ratings of the separate components of the pair.

In an earlier study (Corah, Feldman, Cohen, Gruen, Meadow, & Ringwall, 1958) of part of the EPPS, it was found that for the majority of pairs Ss could readily choose one member of the pair as being more desirable. Further, there was a high correlation between SD ratings of the pairs and degree of endorsement of the items within the pairs. A replication of this study on the entire EPPS (Edwards, Wright, & Lunneborg, 1959) produced a somewhat lower degree of correlation between SD ratings and item endorsement. However, they found an average correlation of around .60 which is still substantial and considerably higher than the .40 correlation reported earlier by Edwards (1959) which was based upon responses to single items rather than pairs. We suggested

that single statements may acquire contextual meaning when paired; hence their SD values may be somewhat altered. An alternative possibility is that item pairs on the EPPS have not been matched adequately. Edwards (1957) noted the difficulty of securing exact matches, for although he tried to match pairs with a difference no greater than .50 of a scale unit (scale units being determined by the method of successive intervals) about 25% of the pairs on the EPPS differed by more than this amount. It is possible that more stringent criteria for matching pairs might indeed minimize the influence of SD. On the other hand, it is also possible that the forced choice method itself produces inherent difficulties. The present study was conducted to elucidate these two possibilities.

PROCEDURE

Items pertaining to achievement, hostility, and sex were culled from three sources: (a) the EPPS, (b) the Minnesota Multiphasic Personality Inventory (MMPI), and (c) the Buss and Durkee (1957) hostility inventory. The writers also devised additional items. This preliminary harvest yielded 132 items written as first person statements.

To insure the face validity of the items, three experienced clinical psychologists independently rated each item for its appropriate need category. A doubtful category was included for items which did not fall readily into the three categories. Items were retained for further use only if all three judges agreed on the need category being represented or if two agreed and the third placed the item in the doubtful category. A pool of 126 items was retained including 30 achievement, 66 hostility, and 30 sexual items.

The 126 items were randomly mixed in the usual personality questionnaire form. The questionnaire was then administered to 100 male students taking introductory psychology at the University of Buffalo. They were instructed to rate each separate item on a nine-interval scale ranging from extreme social undesirability to extreme social desirability. To facilitate the next step, that is, matching items for social desirability, the items were scaled by the method of successive intervals (Edwards, 1952). The

¹ Now at Washington University Medical School.

procedure to this point was essentially the same as the method used by Edwards in constructing the EPPS.

On the basis of matching scale values, 36 item pairs were formed.² Because our aim was to match as closely as possible, a stringent criterion of .10 of a scale unit was adopted for matching which contrasts with Edwards use of .50 of a scale unit. Actually it was possible to comprise 30 of the item pairs with a discrepancy of only .05 or less of a scale unit. It may be of interest to note that the achievement items had the widest range of SD ratings while few hostility items were rated extremely favorable and few sex items extremely unfavorable. Nevertheless, in matching pairs, an attempt was made to cover as wide a range of social desirability as this particular pool of items permitted. The 36 item pairs were composed as follows:

6 ach-ach	6 ach-host
6 host-host	6 ach-sex
6 sex-sex	6 host-sex

In each set of six pairs combining two needs, one need was presented in the first position in three pairs and in the second position in the other three pairs.

The 36 item pairs, arranged in random order, were administered to a new group of 98 male students. The test instructions were as follows:

Here are 36 pairs of statements. Select the statement in each pair which you think is more socially desirable, that is, the statement that would make a person look better to other people if it were said of him.

RESULTS

If SD was controlled by the matching, then within the limits of sampling errors, each member of a pair would be chosen with equal frequency ($p = .50$) as being more desirable. The binomial expansion was used as a basis for testing this proposition. Before considering the main results, we tested for the presence of a different kind of response set, namely a systematic position preference for either the first or second member of the pair. The findings were negative; hence we can conclude that position response sets were negligible in determining responses to these item pairs.

Turning to the main results, Table 1 shows that about half the pairs (19 of 36) yielded significant differences in SD ratings favoring one member of the pair. This number of significant differences is far beyond chance ex-

TABLE 1

NUMBER OF PAIRS SHOWING SIGNIFICANT DIFFERENCES IN SOCIAL DESIRABILITY RATINGS

Paired needs	No. of significant differences within each set of six Pairs (.05 level)	A more favorable than B	B more favorable than A
A.ach-B.ach	3		
A.host-B.host	1		
A.sex-B.sex	3		
A.ach-B.host	5	3	2
A.ach-B.sex	4	4	0
A.host-B.sex	3	2	1

pectations. It is evident that our stringent matching procedures failed to equate items for SD. This is further illustrated by the fact that the percentage of Ss choosing the A member of each pair as being more desirable ranged from 12 to 85%. Even items paired within the same need category, e.g., ach-ach pairs, were evaluated differentially. On the basis of the results of our previous study (Corah et al., 1958), we had anticipated a tendency for achievement items to be rated more favorably when paired with the other needs, but as can be seen this only held for the ach-sex pairings.

The 36 pairs were grouped into the 12 pairs with the lowest (most undesirable), 12 with the middle, and 12 with the highest SD ratings. We then simply counted the number of significant pair differences occurring within each zone which turned out to be 9, 2, and 8 for low, middle, and high, respectively. Thus it is apparent that the pairs were equated relatively better for SD in the middle range than at either extreme.

DISCUSSION

The present findings confirm the results of our previous study in suggesting that social desirability is not readily minimized by the use of the forced choice format even with extremely careful matching. Perhaps the method itself bears closer scrutiny. It seems possible that the paired format invites a comparison between the components of the pair which leads to finer evaluative distinctions than

² Copies of either the 126-item questionnaire or the 36 paired items may be obtained on request.

would be elicited in responding to single statements. In this respect, the forced choice method may actually heighten the problem of control for social desirability.

It would seem to follow that more attention must be paid to item matchings if the forced choice method is to realize the objective of controlling for SD. If the test constructor bases his matchings on SD ratings of single statements, then our results suggest the greater likelihood of equating pairs drawn from the middle range of SD. Such a restriction on item selection, however, poses serious limitations for many purposes when extreme statements might be more appropriate. We had suggested that:

The method of successive intervals to scale items achieved only a first approximation of equal pairings, and that additional judgments of the item pairs themselves are necessary along with revisions in pairing before the variable of social desirability can be eliminated from the EPPS (Corah et al., 1958).

This method of successive approximation was used by Gross (1959) in constructing a forced choice scale measuring inner-other direction. Statistical analysis revealed that the resulting item pairs were equated for SD. The method is tedious but Gross' work demonstrates that greater control of SD can be achieved.

Our research is certainly consistent with Edwards' findings (1957) that SD is a pervasive factor. It is not surprising that verbal tests are markedly affected by such test taking attitudes. Along related lines, Osgoods' work (1952) points up the fact that the connotative meaning of words derives in good measure from an evaluation factor, that is, people can hardly use words without assigning an evaluative meaning to them. We would argue that the forced choice format offers possibilities of controlling SD, but this form of inventory may heighten awareness of the comparative desirability of the component members of a pair; hence, control of SD requires special care in matching pairs.

SUMMARY

The present study was a careful attempt to explore the adequacy of forced choice item pairs in eliminating the factor of social de-

sirability. Items related to the needs of achievement, sex, and hostility were rated separately on a nine-interval SD scale by 100 men. Items were then paired both within and between need areas on the basis of their scale values. Stringent criteria for matching were used. The 36 pairs obtained were given to another group of 98 men to see if they could distinguish between the members of each pair on the basis of SD. About one-half of the pairs yielded significant differences in these judgments. The items of pairs with low and high SD values appeared to be more easily distinguished than those with middle values. It was concluded that carefully matched items in a forced choice format do not readily minimize social desirability and that the forced choice method may actually heighten an S's ability to make discriminations on the basis of this factor.

REFERENCES

- BUSS, A. H., & DURKEE, A. An inventory for assessing different kinds of hostility. *J. consult. Psychol.*, 1957, 21, 343-349.
- CORAH, N. L., FELDMAN, M. J., COHEN, I. S., GRUEN, W., MEADOW, A., & RINGWALL, E. A. Social desirability as a variable in the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1958, 22, 70-72.
- CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, 6, 475-494.
- EDWARDS, A. L. The scaling of stimuli by the method of successive intervals. *J. appl. Psychol.*, 1952, 36, 118-122.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- EDWARDS, A. L. Revised manual for the Edwards Personal Preference Schedule. New York: Psychological Corp., 1959.
- EDWARDS, A. L., WRIGHT, C. E., & LUNNEBORG, C. E. A note on "Social desirability as a variable in the Edwards Personal Preference Schedule." *J. consult. Psychol.*, 1959, 23, 558.
- GROSS, H. W. The relationship between insecurity, self-acceptance, other-direction, and conformity under conditions of differential social pressure. Unpublished doctoral dissertation, University of Buffalo, 1959.
- MEEHL, P. E., & HATHAWAY, S. R. The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, 30, 525-564.
- OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, 49, 197-237.

(Received November 2, 1959)

THE CONNOTATIVE MEANING OF SEXUAL SYMBOLS

ROSALEA A. SCHONBAR AND JOEL R. DAVITZ

Teachers College, Columbia University

According to Freud (1953a, p. 354), symbols derive meaning from their similarity in form to the male or female genitalia; pointed or elongated objects are therefore male symbols, while round, containing objects are female symbols. The meaning, according to Freud, is inherent in the structure of the object itself, and therefore, at least by implication, transcends cultural differences.

Although Freud's theory forms the basis of assumptions frequently made in the interpretation of dreams, projective test responses, and artistic productions, recent studies present conflicting evidence concerning its validity. The findings of Starer (1955), Jones (1956), and Stennett and Thurlow (1958) support the Freudian hypothesis, while those of Levy (1954) and Barker (1957) do not.

In previous studies, the experimental procedures involved some variation of matching elongated or rounded objects with a masculine or feminine designation. In general, the methods used may be grouped into several levels of directness. The matching of objects with male or female names (Levy, 1954; Starer, 1955; Stennett & Thurlow, 1958) is the most direct, and, as Levy recognized, the most open to conscious control or to defensive behavior. Jones (1956) and Barker (1957) used similar approaches, asking subjects (Ss) to identify the object as male or female in personality, or to make it a masculine or feminine character in a children's book. While, in itself, this method is still fairly direct, Barker's introduction of a reaction time measure served as a check on blocking or avoidance responses. Levy also utilized a paired-associates learning task, perhaps the most indirect of the methods.

Nevertheless, the basic task remains the same, one of "naming" or classifying, and

there is considerable doubt that this denotative function is the most appropriate for the testing of the Freudian hypothesis. One of the most important elements of symbolic meaning remains untouched, namely the connotative, or "feeling tone," aspect of the symbol. In clinical use, it is this aspect which is the most significant. The meaningfulness of a symbol, sexual or otherwise, lies in the emotional atmosphere it creates. It is therefore the purpose of this study to test the validity of the Freudian theory of symbols in terms of connotative rather than denotative meaning.

Barker's research (1957), upon which the present study is based, introduced the cultural meaning of symbols as a major variable. She showed her Ss drawings of objects which could be classified as male or female according to Freudian theory or to cultural use. The stimuli were classified into four categories: (a) those with Freudian sexual designations, but low in cultural sexual meaning (i.e., snail, fish); (b) those for which Freudian designations and cultural associations were in agreement (i.e., purse, gun); (c) those for which Freudian and cultural sexual designations were in conflict (i.e., needle, drum); and (d) those which had neither Freudian nor cultural sexual meaning (i.e., book, table). Barker found that, where culture was not a variable, sexual designations did not differ from chance; where culture was a variable, sexual designation was in terms of cultural meaning, whether it agreed with or conflicted with the Freudian prediction.

The hypothesis to be tested in the present research is as follows: Regardless of the degree to which they conform to cultural associations, pointed or elongated objects are designated as male, while rounded, containing objects are designated as female.

In addition, since the two studies with negative findings (Barker, 1957; Levy, 1954) used children as Ss, while the other three used young adults, Stennett and Thurlow (1958) have suggested that age may be the significant variable. The present research is designed to cast light on this issue.

METHOD

The Ss of this research were 49 undergraduate students enrolled in an introductory psychology course at Long Island University. There were 17 males and 32 females.

Both denotative and connotative designations of the sexual meaning of each symbol were obtained. The denotative designation was obtained by asking Ss to indicate whether each of the 24 symbols would be most appropriate as a male figure or a female figure if it were used in a children's storybook. The connotative designations were obtained by asking Ss to describe the 24 symbols plus a male and a female figure on 12 scales of the semantic differential.

To determine whether an S described an object on the semantic differential as male or female, difference scores were computed between the S's ratings of each symbol and his ratings of the male and female figures. If the difference in ratings over the 12 scales between the symbol and the male figure was smaller than the difference between the symbol and the female figure, that symbol was designated as male. Similarly, if the difference between the symbol and the female figure was smaller than the difference between the symbol and the male figure, that symbol was designated as female.

In all cases, line drawings of the symbols were presented to the Ss, and they were asked either to designate the symbol drawn as male or female or to describe it using the 12 rating scales. All Ss first rated the symbols on the semantic differential and then made the denotative designations.

Of the 24 symbols used in the present study, 23 were obtained from Barker's study (1957), using the designations based on psychologists' judgments of Freudian and cultural meanings. One symbol, tobacco jar, was added for the present study and was designated as female according to Freudian theory and male according to cultural theories.

The 24 symbols were divided into four groups, six symbols in each group. Group 1 consisted of six symbols with a clear Freudian meaning but without an agreed upon cultural designation. These were, according to Freudian theory, three female symbols—snail, suitcase, apple—and three male symbols—fish, ice cream cone, 3. Group 2 consisted of six symbols about which the Freudian and cultural theories agreed. These were three male symbols—hammer, airplane, gun—and three female symbols—stove, vase, hatbox. Group 3 consisted of three symbols whose symbolic sexual meaning according to Freudian theory was male, but whose cultural associa-

tions were female. These were: needle, broom, lipstick. The other three symbols in Group 3 were those whose symbolic sexual meaning according to Freudian theory was female, but whose cultural associations were male. These were: tobacco jar, basketball net, drum. Thus, Group 3 comprised symbols for which the Freudian and cultural theories were opposed in designation of symbolic sexual meanings. Group 4 consisted of six symbols about which there was neither a clear-cut Freudian nor cultural meaning. These were: leaf, table, 9, chair, book, glasses.

RESULTS

Two sets of data were obtained: the denotative designations of either male or female for each symbol and the connotative responses inferred from the semantic differential ratings. For each symbol, the binomial probability was computed using the normal curve approximation. Using a two-tailed test, at least 31.4 of the choices must be either male or female to achieve significance at the .05 level. The data and results of the analysis are summarized in Tables 1 and 2.

Denotative Designation

Table 1 summarizes the results concerning the denotative designations of the 24 symbols. Group 1 contained those symbols for which there was a Freudian designation but no clear-cut cultural designation of sexual meaning. As indicated in Table 1, no consistent pattern of results emerged from these data. For the three symbols designated in Freudian theory as feminine, snail and suitcase were chosen as masculine, and apple was chosen as feminine. For the three symbols designated as masculine in Freudian theory, the distribution for 3 did not differ significantly from chance, ice cream cone was feminine, and fish masculine.

For Group 2, which comprised those symbols for which Freudian and cultural theories were in agreement, the results were in accord with theoretical expectation. Hammer, airplane, and gun were designated as masculine, and stove, vase, and hatbox were feminine. In Group 3, which contained the symbols about which Freudian and cultural theories were in opposition, the results clearly supported the cultural theories. Needle, broom, and lipstick were designated as feminine, and tobacco jar, basketball net, and drum were

TABLE 1
DENOTATIVE DESIGNATIONS
Number of Subjects

Symbol	Male Ss		Female Ss		Total		<i>p</i> ^a
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	
Group 1							
Snail	10	7	25	7	35	14	.05
Suitcase	11	6	20	12	31	18	<i>ns</i>
Apple	10	7	9	23	19	30	<i>ns</i>
3	12	5	18	14	30	19	<i>ns</i>
Ice cream cone	12	5	1	31	13	36	.05
Fish	13	4	27	5	40	9	.05
Group 2							
Hammer	17	0	32	0	49	0	.05
Airplane	17	0	30	2	47	2	.05
Gun	17	0	32	0	49	0	.05
Stove	0	17	2	30	2	47	.05
Vase	7	10	10	22	17	32	.05
Hatbox	0	17	0	32	0	49	.05
Group 3							
Needle	2	15	5	27	7	42	.05
Broom	1	16	1	31	2	47	.05
Lipstick	3	14	1	31	4	45	.05
Tobacco jar	17	0	32	0	49	0	.05
Basketball net	16	1	32	0	48	1	.05
Drum	17	0	32	0	49	0	.05
Group 4							
Leaf	8	9	15	17	23	26	<i>ns</i>
Table	8	9	17	15	25	24	<i>ns</i>
9	12	5	18	14	30	19	<i>ns</i>
Chair	10	7	18	14	28	21	<i>ns</i>
Book	13	4	9	23	22	27	<i>ns</i>
Glasses	4	13	18	14	22	27	<i>ns</i>

^a Probabilities computed on basis of total distribution.

masculine. These designations, of course, are in accord with the cultural associations of the symbols rather than the form of the object. In Group 4, for which there was neither a Freudian nor cultural designation, the distribution of masculine and feminine designations did not differ from chance for any of the symbols.

These data are in complete agreement with the results reported by Barker (1957). In the denotative designation of the sexual meaning of symbols, the designation clearly is determined by the cultural association of the object rather than by its form.

Connotative Designation

Table 2 summarizes the data obtained from the inferred designations based on the semantic differential descriptions. Group 1 contained those symbols for which there was a Freudian but no culturally relevant sexual meaning. The distribution for only one sym-

bol attained significance, ice cream cone as feminine, which contradicted Freudian theory.

In Group 2, the three symbols, hammer, airplane, and gun, which both Freudian and cultural theories designate as masculine were indeed described as masculine by the Ss in this study. However, the symbols theoretically designated as feminine revealed contradictory results: stove was described as masculine, hatbox as feminine, and the distribution for vase did not differ significantly from chance.

In Group 3, the distributions for needle and broom did not differ significantly from chance, while lipstick was described as feminine, in contradiction to Freudian theory. Tobacco jar was described as feminine, which supports the Freudian position, but both basketball net and drum were described as masculine, which supports cultural theories. Thus, for those symbols about which Freudian and cultural theories are in opposition, no clear-cut pattern emerged.

As would be expected on the basis of both Freudian and cultural points of view, the symbols in Group 4, which had no clear theoretical designation, showed no consistent pattern of results. Leaf was described as feminine, and table and chair as masculine, but the distributions for 9, book, and glasses did not differ significantly from chance.

In summary, the connotative designations obtained on the basis of semantic differential descriptions did not reveal as clear a pattern of results as did the direct designations. However, the data tend to contradict rather than support the Freudian position. Of the six symbols about which there was theoretical opposition, the results for three symbols supported the cultural point of view, while results for only one symbol were in the direction predicted by the Freudian position. When both Freudian and cultural theories were in agreement that the symbols were masculine, as in Group 2, the results were in substantial agreement with both theories. But when there was a Freudian designation with no particular cultural designation, the only distribution which differed significantly from chance was in the direction opposite to that predicted by Freudian theory.

TABLE 2
CONNOTATIVE DESIGNATIONS
Number of Subjects

Symbol	Male Ss		Female Ss		Total Ss		<i>p</i> ^a
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	
Group 1							
Snail	10.5 ^b	6.5	17	15	27.5	21.5	<i>ns</i>
Suitcase	12	5	18.5	13.5	30.5	18.5	<i>ns</i>
Apple	5.5	11.5	17.5	14.5	23	26	<i>ns</i>
3	7.5	9.5	12	20	19.5	29.5	<i>ns</i>
Ice cream cone	1.5	15.5	10	22	11.5	37.5	.05
Fish	7	10	12	20	19	30	<i>ns</i>
Group 2							
Hammer	13	4	28.5	3.5	41.5	7.5	.05
Airplane	15	2	26.5	5.5	41.5	7.5	.05
Gun	15	2	20	12	35	14	.05
Stove	15	2	23	9	38	11	.05
Vase	9.5	7.5	15	17	24.5	24.5	<i>ns</i>
Hatbox	3.5	13.5	12	20	15.5	33.5	.05
Group 3							
Needle	4	13	15	17	19	30	<i>ns</i>
Broom	9	8	11.5	20.5	20.5	28.5	<i>ns</i>
Lipstick	6	11	10	22	16	33	.05
Tobacco jar	5	12	10	22	15	34	.05
Basketball net	15	2	18	14	33	16	.05
Drum	9	8	22.5	9.5	31.5	17.5	.05
Group 4							
Leaf	2	15	11.5	20.5	13.5	35.5	.05
Table	12	5	20.5	11.5	32.5	16.5	.05
9	8.5	8.5	22.5	9.5	31	18	<i>ns</i>
Chair	11	5	21	11	33	16	.05
Book	11.5	5.5	13	19	24.5	24.5	<i>ns</i>
Glasses	7	10	16	16	23	26	<i>ns</i>

^a Probabilities computed on basis of *t*-test.

^a Probabilities computed on basis of totals.

^b In cases of ties, 0.5 was credited to each designation.

DISCUSSION

The results of this study do not support the Freudian theory of the meaning of symbols. Where culture was not a factor, there was no clear-cut definition of sexual meaning; where culture was a factor, it was, in general, culture rather than form which determined the sexual meaning of the object. This was true for both the denotative and connotative designations of sexual meaning.

Freud's (1953b) own clinical use of symbols was characterized by interpretation based upon a combination of the patient's individual associations to objects in dreams and the ap-

plication of the theory tested here. It is clear that, if cultural factors determine sexual meaning, universality of meaning can exist only to the extent that cultural elements are similar or identical for large numbers of people. To understand sexual symbols for a given patient, therefore, individual associations may be made more meaningful by an understanding of the relationship of the symbol to sex-role expectations in his particular cultural background. This view is in essential agreement with Fromm's (1951) statements concerning dream interpretation.

The conformity of sexual meaning to cul-

tural expectation was clearer on the direct, or denotative, designations than on the inferred, or connotative, designations. This may be due, in part, to the relative lower reliability of scores derived by this method, or it may be due to the complexity of factors contributing to connotative responses. Moreover, the cultural sexual designation presumably was based on associations related to sex role; to the degree that the sex-role associations may have varied in clarity for the stimulus objects, variability in designation might be expected. In this regard, it is interesting that culturally female objects are described as such somewhat less consistently and strongly than culturally male symbols are designated as male.

The clearest inferred designation of sexual meaning occurred when form and culture coincided on masculine objects. This is related, in part, to the fact that masculine items were identified more consistently than feminine items. But, unless this is a chance result, this is the only indication in our findings that form plays any role in sexual meaning. It is possible that a complex function operates, wherein the Freudian interpretation, as a part of modern culture, may operate to an unknown degree under certain unspecifiable conditions. Where form was the only determinant, sexual meaning was indeterminate. In the studies which have supported the Freudian view, the stimuli were abstract forms rather than familiar objects, thus having no specific cultural meaning. It is possible that, under these conditions, Ss utilized whatever associations were available to them, and that either similarity to sex organs as such or the knowledge that such a theory exists determined the responses, at least under direct testing. This is of interest, of course, as is the question of what other conditions might lead to similar determinations of sexual meaning; on the other hand, when dreams or responses to projective tests are being interpreted, the symbols involved are

almost always objects rather than geometric forms.

Inspection of our results reveals that men and women did not differ in their connotative reactions to the objects. Moreover, the repetition of Barker's method (1957) with young adults resulted in a replication of the results she obtained with children. It is, therefore, not likely that age is an important factor.

SUMMARY

The Freudian hypothesis that pointed or elongated objects are male symbols, while round, containing objects are female symbols was tested using both denotative and connotative responses. For both types of designations, the results did not support the Freudian theory. In general, culture rather than form determined the sexual meaning of the objects. These results were discussed in regard to clinical interpretation of the meaning of symbols and in relation to previous research.

REFERENCES

- BARKER, ELIZABETH D. The perception of sexual symbolism. Unpublished doctoral dissertation, Columbia University, 1957.
- FREUD, S. The interpretation of dreams (1900). In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud*. Vols. 4 & 5. London: Hogarth, 1953. Pp. 1-686. (a)
- FREUD, S. Fragment of an analysis of a case of hysteria (1905). In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud*. Vol. 7. London: Hogarth, 1953. Pp. 3-122. (b)
- FROMM, E. *The forgotten language*. New York: Rinehart, 1951.
- JONES, A. Sexual symbolism and the variables of sex and personality interpretation. *J. abnorm. soc. Psychol.*, 1956, 53, 187-190.
- LEVY, L. H. Sexual symbolism: A validity study. *J. consult. Psychol.*, 1954, 18, 43-46.
- STARER, E. Cultural symbolism: A validity study. *J. consult. Psychol.*, 1955, 19, 453-454.
- STENNETT, R. G., & THURLOW, M. Cultural symbolism: The age variable. *J. consult. Psychol.*, 1958, 22, 496.

(Received November 19, 1959)

COMMUNICATED ANXIETY IN A TWO-PERSON SITUATION¹

PATRIK O. MATTSSON

New Jersey State Hospital at Ancora

The interpersonal basis of anxiety has long been recognized in psychological theory. Among others, Sullivan (1953), Fromm (1941), and Horney (1937) have proposed theories based upon interpersonal relations.

Sullivan (1953) described the process underlying the interpersonal communication of anxiety in terms of "empathy" (p. 41). Sullivan's (1947) writing indicates that "euphoria" or the relief of anxiety can similarly be conveyed through empathy. While he considered empathy to be most applicable to describe the arousal of anxiety in the infant by the mother, he also assumed that this mode of emotional communication can last throughout life. Although he pointed out that interpersonal communication of anxiety is generally recognized in psychological practice, Sullivan did not offer further interpretation of the nature of empathy.

While previous experimental studies have not directly investigated interpersonal communication of anxiety between two physically present persons, variables relevant to anxiety in an interpersonal situation have been considered. Reiser, Reeves, and Armington (1955) demonstrated that release of emotional tension through verbalization depended on the social relationship of the persons interacting. Seidman, Bensen, Miller, and Mee-land (1957) found that subjects (Ss) had greater tolerance for self-administered electric shock when together with another person

who they thought shared the shock. Ruesch and Prestwood (1949) found that listening to voice and speech recordings of anxious patients in psychotherapy induced anxiety in the listeners.

A number of studies have approached the problem of communicated emotions through animal experimentation. Masserman (1943), Liddell (1950), Davitz and Mason (1955), and Lirzman (1955) demonstrated that the presence of a nonfearful animal had a calming effect on a fearful one.

The purpose of this study was an experimental investigation of increase and reduction of anxiety in an interpersonal setting. The following hypotheses were tested:

1. A person who interacts with a more anxious person will himself become more anxious.
2. A person who interacts with a less anxious person will himself become less anxious.

DESIGN

The experimental procedure involved the interaction of pairs of Ss in a cooperative task situation. The setting in which interaction took place was designed to promote maximum opportunity for communication, both verbal and nonverbal. Anxiety was experimentally induced by means of electric shock. Changes in Ss' anxiety were measured through physiological and psychological manifestations.

Subjects

The Ss were 80 undergraduate and graduate male students at Columbia University. Median age was 20, and the interquartile range was 19 to 24. The Ss who worked together did not know each other previously. In order to obtain at least a rough estimate of the S's chronic level of anxiety, a short form of the Taylor Manifest Anxiety Scale (MAS) was administered (Bendig, 1956). The groups of Ss compared for the purpose of testing the hypotheses were not significantly different from each other ($p > .10$) in terms of anxiety as measured by this scale.

¹ The present paper is part of a dissertation presented in partial fulfillment of the requirement for the degree of Doctor of Philosophy at Teachers College, Columbia University, 1959. The writer wishes to express his thanks to Joel R. Davitz (doctoral committee chairman), Kenneth F. Herrold, Rosalea A. Schonbar, and Laurance F. Shaffer for their invaluable advice and criticism.

The Ss were randomly assigned to one of the following kinds of experimental situations, each employing 20 Ss: nonshocked *S* working with shocked *S*, NS(wS); shocked *S* working with nonshocked *S*, S(wNS); nonshocked *S* working with another nonshocked *S*, NS(wNS); shocked *S* working with another shocked *S*, S(wS).

Experimental Induction of Anxiety

Low current electric shock is a harmless but rather effective means of raising the anxiety level of an individual (Erickson & Wechsler, 1955). In this study an electrical shock was generated by transformer (inductorium), supplied by a source of 3-, 6-, or 12-volt direct current. The intensity of the output (electric shock) could be varied continuously from 50 to 350 volts. Electrodes, to be attached to Ss' wrists, were designed so that one person alone or two persons simultaneously could get shock. Stimulations were set off by a silent switch, operated by the experimenter (*E*), and were given in a random but predetermined and fixed order for each *S*. The intensity of shock was set according to each *S*'s tolerance for such shock.

The establishment of the tolerance level for shock was done by giving the shocked Ss electric stimulations of low voltage, and then gradually increasing the voltage to a point where the Ss, in their own estimation, considered the shock to be "uncomfortable." The level of voltage used during the actual experiment was twice that which the Ss had responded to as uncomfortable.

Measures of Anxiety

This study evaluated change in anxiety through concomitant tension changes. Physiological tension was measured by changes in systolic blood pressure (BP) and by changes in finger perspiration. Emotional tension was measured by changes in response to a self-rating scale.

BP was measured with a sphygmomanometer with the unit of measure millimeter mercury column. Increased blood pressure was seen as an indication of increased tension (anxiety). Each of the measures of blood pressure used in the analysis of data was based upon the mean of two estimates obtained at each occasion of measurement. Based upon 25 of such paired estimates, selected at random, a reliability coefficient (Spearman rank correlation coefficient) of .99 was obtained.

Finger perspiration was measured by a colorimetric method similar to the one described by Mowrer, Light, Luria, and Zeleny (1953). At each measurement the palmar surface of one finger on each hand was painted with a solution of ferric chloride in acetone and allowed to dry. The treated fingers then were pressed for three minutes upon drawing paper which earlier had been impregnated with tannic acid. The chemicals on the finger and the paper and the moisture of the perspiration combined to produce a

dark fingerprint. The more perspiration was present, the darker was the print. The darkness of the print was measured by a photoelectric device (densitometer). The measure thus yielded was the Finger Perspiration Index (FPI), which could take a value from 0 to 40. A decreased score meant increased finger perspiration, i.e., increased tension. The mean value of two fingerprints obtained at each measurement was taken as the FPI for the *S* at the time of the measurement. Based upon 25 of these paired estimates, a reliability coefficient (Spearman) of .90 was obtained.

Psychological manifestations of anxiety were measured by a self-rating scale, specifically designed to reflect situational tension changes during the experimental procedure. The scale consisted of six subscales, each containing common verbal expressions of emotional tension, arranged on a continuum from one extreme to the other along a 150-millimeter long six-point scale. The six scales included: uneasy-at ease, calm-restless, undisturbed-disturbed, comfortable-uncomfortable, upset-peaceful, and relaxed-tense.

The scales were presented to the Ss in the form of a booklet with one scale *per* sheet. At repeated administrations the same scales were presented to the Ss again, but arranged in a different order. When scoring the questionnaire, the distance in millimeters from the "low tension" end of each scale to the place where the *S* had marked the scale with a checkmark was measured. An *S*'s rating scale score (*R*) was the mean distance from the low tension end to the checkmark for the six scales. The *R* could yield a value from 0 to 150. An increase in score was seen as an indication of increased tension (anxiety). The composite score for three of the scales was correlated with the score for the three remaining scales, yielding a measure of internal consistency. Based upon a random sample of 25 rating scales, a reliability coefficient (Spearman) of .88 was obtained.

Task

As a means of facilitating, as well as controlling, the interaction of the Ss within each pair, a simple work task was given. The task material was two different picture puzzles, each containing 27 pieces, having a finished size of 10 × 12 inches (Task I and Task II).

The rules for putting the puzzles together were formulated so that each member of the cooperating pair was dependent on the other member for any and all progress in the task. The purpose was to make the partners "significant" in each others' eyes, at least as far as working with the task was concerned. The rules also contained specifications that required the members to look at each other, talk and listen to each other, maximizing the opportunity for verbal and nonverbal communication.

For each pair, as a unit, a measure of work was obtained by counting how many pieces had been placed correctly before the task was discontinued. A

correlation (Spearman) between scores for Task I and Task II, based upon 30 pairs, selected at random, was .34 ($p < .05$).

Procedure

At the enlistment for the experiment, each *S* was told that he was to participate in a study on interpersonal relations. During the session the *Ss* were seated at a table facing each other. The *E* was seated at one end of the same table.

Because the *E* was present with the *Ss*, it may appear that the experiment was dealing with a three-person relationship rather than with a two-person interaction. However, the factor of the *E* was constant, and even though a shocked *S* might see the *E* in a different role than a nonshocked *S*, the experimental design and measurements were such that the *E*'s presence would only influence the results minimally.

The pairs of *Ss* were divided into three groups, for which the procedure was somewhat different: NS(wS), S(wS), and NS(wNS). The procedure for the pairs of *Ss* in which one member received shock but the other member did not—S(wNS)—was as follows:

Step 1. Measurement of finger perspiration and blood pressure. Administration of self-rating scale ("Before" measurement).

Step 2. Electrodes attached to shocked *S*. Shock tolerance determined.

Step 3. Cooperative work with puzzle (Task I). Electric shock given to shocked *S*. Time: 5 minutes.

Step 4. Measurements repeated ("Middle" measurement).

Step 5. *Ss* work with Task II, the other puzzle. Electric shock given to shocked *S*. Time: 5 minutes.

Step 6. Measurements repeated ("After" measurement).

Step 7. Electrodes removed. Additional rating scales regarding the *Ss*' evaluation of each other and the short form of the MAS given.

It is to be noted that the instructions to the *Ss* regarding the electric shock procedure included an emphasized statement that the partners of the shocked *Ss* were *not* to receive shock at any time during the experiment. Such instruction was intended to insure, within reason, that hypothesized increased anxiety in the nonshocked *Ss* was not to be a function of possible anticipation that they too were to receive shock some time during the experiment. This cautionary procedure seems to have been generally reassuring to the nonshocked *Ss*, as evaluated during informal questioning after the experiment.

The procedure for the pairs where each member received shock was the same as the above procedure, except that the steps dealing with electric shock were applied to both *Ss*. For the pairs where neither *S* received shock, the steps relating to electric shock were omitted.

In summary: The three occasions during the experiment when the three measures of tension were

administered are referred to as the Before, Middle, and After measurements. The time elapsed between the Before and Middle measurements was approximately 16 minutes for pairs where no shock was administered, and 20 minutes where both *Ss* got shock. The time between the Before and After measurements was twice as long. The measures of tension were: BP, FPI, and response to a rating scale. The measure of chronic anxiety level was a short form of the MAS. Additional measures, in the form of two graphic rating scales, expressed the *Ss*' evaluation of each other during the experiment.

Treatment of Data

The analyses of the results were based upon the difference between the scores at the Before and Middle measures as well as the difference between scores at the Before and After measures. The difference scores represented the *Ss*' changes in tension during the time interval between the occasions of measurement.

Each group of *Ss* provided two distributions of difference scores for each measure of anxiety. The hypotheses were tested by comparing the distribution of difference scores for appropriate groups of *Ss*. The .05 level of significance was used to determine the acceptance or rejection of the hypotheses tested. Because no assumptions were made regarding the shapes of the populations from which the samples were drawn, nonparametric tests were used throughout the study. The hypotheses were tested by the Mann-Whitney *U* test. One-tailed tests of significance were used to test the hypotheses, while two-tailed tests were used for additional analyses or testing relationships where the directions of change, if any, were not anticipated.

TABLE 1
DIFFERENCE SCORES AND VALUES OF *U* FOR
THE FINGER PERSPIRATION INDEX

Time Period	Groups Compared		Value of <i>U</i> *
	Nonshocked (with shocked)	Nonshocked (with nonshocked)	
Before to Middle	<i>Q</i> ₁ = - 5.0 ^a <i>Md</i> = - 1.2 <i>Q</i> ₃ = 2.3	<i>Q</i> ₁ = - 1.7 <i>Md</i> = 0.2 <i>Q</i> ₃ = 2.4	194.5
Before to After	<i>Q</i> ₁ = - 9.3 <i>Md</i> = - 1.1 <i>Q</i> ₃ = 0.3	<i>Q</i> ₁ = - 3.1 <i>Md</i> = 0.3 <i>Q</i> ₃ = 3.4	96.5

Note.—*Q*₁, *Md*, and *Q*₃ refer to the First Quartile, Median, and Third Quartile of the distribution of difference scores.

* A negative difference score for FPI indicates increased finger sweat.

* Significance at the .05 level = 138, and at the .01 level = 114.

TABLE 2

DIFFERENCE SCORES AND VALUES OF U FOR THE BLOOD PRESSURE

Time Period	Groups Compared		Value of U^*
	Nonshocked (with shocked)	Nonshocked (with nonshocked)	
Before to Middle	$Q_1 = -8$ $Md = 3$ $Q_3 = 6$	$Q_1 = -6$ $Md = -4$ $Q_3 = -1$	135.5
Before to After	$Q_1 = -9$ $Md = -2$ $Q_3 = 4$	$Q_1 = -12$ $Md = -7$ $Q_3 = -3$	131

* Significance at the .05 level = 138, and at the .01 level = 114.

RESULTS

Hypothesis 1

Rephrased in operational terms, the first hypothesis stated that in a cooperative task situation, nonshocked Ss who work with nonshocked Ss show greater anxiety than nonshocked Ss who work with other nonshocked Ss. Tables 1, 2, and 3 show the comparison of the two groups of nonshocked Ss: NS(wS) and NS(wNS).

The hypothesis was tested twice with each measure of anxiety. As indicated in Table 1, the hypothesis was supported for the Before to After FPIs. Table 2 indicates support of

TABLE 3

DIFFERENCE SCORES AND VALUES OF U FOR THE RATING SCALE SCORE

Time Period	Groups Compared		Value of U^*
	Nonshocked (with shocked)	Nonshocked (with nonshocked)	
Before to Middle	$Q_1 = -5$ $Md = 7$ $Q_3 = 38$	$Q_1 = -14$ $Md = 2$ $Q_3 = 9$	137.5
Before to After	$Q_1 = -15$ $Md = -6$ $Q_3 = 1$	$Q_1 = -18$ $Md = -8$ $Q_3 = 8$	194

* Significance at the .05 level = 138, and at the .01 level = 114.

TABLE 4

DIFFERENCE SCORES AND VALUES OF U FOR THE FINGER PERSPIRATION INDEX

Time Period	Groups Compared		Value of U^*
	Shocked (with shocked)	Shocked (with nonshocked)	
Before to Middle	$Q_1 = -5.2$ $Md = -1.3$ $Q_3 = 3.5$	$Q_1 = -5.2$ $Md = -1.8$ $Q_3 = 0.6$	182
Before to After	$Q_1 = -4.9$ $Md = 0$ $Q_3 = 2.1$	$Q_1 = -9.5$ $Md = -5.5$ $Q_3 = 0$	152.5

* Significance at the .05 level = 138, and at the .01 level = 114.

the hypothesis for both the Before to Middle and Before to After BPs. Table 3, dealing with the result of the rating scale, indicates support for the hypothesis for the Before to Middle measure.

Hypothesis 2

Rephrased in operational terms, the second hypothesis stated that in a cooperative task situation shocked Ss who work with other shocked Ss show greater anxiety than shocked Ss who work with nonshocked Ss.

To test the second hypothesis, the two groups of shocked Ss were compared: S(wS)

TABLE 5

DIFFERENCE SCORES AND VALUES OF U FOR THE BLOOD PRESSURE

Time Period	Groups Compared		Value of U^*
	Shocked (with shocked)	Shocked (with nonshocked)	
Before to Middle	$Q_1 = -7$ $Md = 0$ $Q_3 = 8$	$Q_1 = -2$ $Md = 3$ $Q_3 = 7$	183
Before to After	$Q_1 = -5$ $Md = 5$ $Q_3 = 7$	$Q_1 = -6$ $Md = -2$ $Q_3 = 9$	193

* Significance at the .05 level = 138, and at the .01 level = 114.

TABLE 6
DIFFERENCE SCORES AND VALUES OF U FOR
THE RATING SCALE SCORE

Time Period	Groups Compared		Value of U^*
	Shocked (with shocked)	Shocked (with nonshocked)	
Before to Middle	$Q_1 = 9$ $Md = 41$ $Q_3 = 51$	$Q_1 = 14$ $Md = 30$ $Q_3 = 43$	180
Before to After	$Q_1 = 21$ $Md = 45$ $Q_3 = 62$	$Q_1 = 23$ $Md = 39$ $Q_3 = 64$	191.5

* Significance at the .05 level = 138, and at the .01 level = 114.

and S(wNS). Tables 4, 5, and 6 present summaries of the analysis of data relevant to this hypothesis. As indicated in these tables, the result did not support the second hypothesis.

Additional Analyses

The additional analyses concerned interrelationships among the three measures of tension, the relationship of shock to task performance, the relationship of shock tolerance to measures of tension, and Ss' evaluations of their partners.

Interrelationships among measures of tension. Table 7 contains the intercorrelations (Spearman's rho) of the three measures of tension (anxiety).

As indicated in Table 7, there was no substantial correlation between the several expressions of anxiety. It is conceivable that the physiological responses changed in opposite directions for different Ss. Similarly, some Ss may have responded psychologically to anxiety by denying it, i.e., rated themselves as less anxious on the questionnaire. An analysis indicated that the NS(wS) group had a significantly larger variability for the BP scores than the NS(wNS), less anxious, group.

Electric shock and task performance. There was no significant difference between the task performances of pairs where shock was administered—(SwS)—and of pairs where no shock was given—(NSwNS).

Shock tolerance and anxiety measures. Statistical analysis revealed that shock tolerance was independent of all measures of anxiety used in this study.

Subjects' evaluations of each other. The groups of Ss, compared for the purpose of testing the hypotheses, were not significantly different in their evaluation of their likes or dislikes for their partners. However, nonshocked Ss felt that nonshocked partners cooperated better than did shocked partners.

TABLE 7
INTERCORRELATION AMONG TENSION MEASURES

Measures Correlated	Time Period	Non-shocked (with Shocked)	Non-shocked (with Non-shocked)	Shocked (with Shocked)	Shocked (with Non-shocked)
Finger Perspiration and Blood Pressure	Before to Middle	.14	.64*	-.10	.20
	Before to After	-.23	.26	-.16	.32
Finger Perspiration and Rating Scale	Before to Middle	.36	.39*	.23	.37
	Before to After	.26	.09	.06	.35
Blood Pressure and Rating Scale	Before to Middle	-.12	.35	.44*	.04
	Before to After	.02	.25	.08	.19

* Significance at the .05 level = .38, and at the .01 level = .53.

DISCUSSION

The present study demonstrated that anxiety in one person in an interpersonal relationship is communicated and results in increased anxiety in another person in the situation. However, the results failed to support the hypothesis that the presence of a person with relatively lower anxiety reduces the level of anxiety in the more anxious person.

Communication of Anxiety

It may be assumed that in the past various peripheral cues have been associated with increased tension or anxiety. Cardiovascular concomitants of anxiety are manifested by blanching, blushing, fainting, and throbbing of arteries close to the skin surface. Sweating is manifested by moisture and a characteristic odor. Motor disruptions related to anxiety may be cued off by tremors, rapid shallow breathing, forced motions, restlessness, and speech disturbances. Psychological concomitants of anxiety may be evidenced by inappropriate behavior or direct verbalization of feelings of discomfort. As these cues are paired with feelings of anxiety or tension, they may become conditioned stimuli for the arousal of anxiety. While the association of signs and feelings of anxiety primarily involves a person's perception of himself, such association may be generalized so that perception of these signs in other persons also elicits anxiety. Thus, perception of another anxious person increases anxiety in the observer by virtue of previous conditioning of observable peripheral cues to anxiety responses.

In the present experiment, the shocked Ss displayed peripheral cues of tension which were observed by the nonshocked Ss. An account of the increased tension in the nonshocked Ss, in terms of conditioning, would provide one basis for explaining the phenomenon of empathy.

Communication of Comfort

Consideration of the nature of comfort cues as well as possible responses to such cues by the shocked Ss may provide an account for the lack of support for the second hypothesis.

Decreased tension or increased feelings of comfort are associated with such cues as di-

minished blanching or blushing, drying up of sweat, increased muscular control, calmer speech, and more normal breathing pattern. In the present study, the nonshocked Ss were to provide the cues of comfort for the shocked Ss. As indicated by the results supporting the first hypothesis, while the nonshocked Ss working with the shocked Ss may initially have displayed less anxiety, they also became more anxious during the experimental sessions. But the fact that nonshocked Ss who worked with shocked Ss became more anxious would in itself not have precluded the possibility of supporting the second hypothesis. While these Ss did become more tense, they did not, in general, become as tense as the shocked Ss. Therefore, there was at least a relative difference in the anxiety levels of these two kinds of Ss. In spite of this, the nonshocked Ss had no "tension reducing" effects on their shocked partners. It is, of course, possible that the relatively lesser anxiety of the nonshocked Ss was less important than the fact that they exhibited some cues of anxiety or anxiety increase. The magnitude of an anxiety cue may be of less significance than the mere presence or absence of such a cue, however minute.

Certain factors tended to maintain the tension of the shocked Ss, in spite of available cues of comfort. The experimental procedure of anxiety induction through electric shock was continued throughout the interaction of the Ss. Many of the shocked Ss expressed the conviction that the discomfort of the shock became increasingly difficult to endure. While a shocked S thus may have clearly perceived the cues of comfort from his nonshocked partner, increased situational stress may have prevented him from feeling any anxiety relief.

Response to stress was most often verbalized by the shocked Ss as extreme discomfort or tension. Some of the Ss did, however, express anger or resentment towards their nonshocked partners as well as towards the E. Because there is a good deal of overlap in the physiological and psychological concomitants of emotions, such anger or resentment would most likely be registered by the employed measures as increased or maintained anxiety.

Hypothesized anxiety decrease of the shocked S in the (SwNS) group was evaluated against the assumed high level of tension maintained

by the Ss in the (SwS) group. Such assumption may, in fact, have been erroneous, contributing to the lack of support for the second hypothesis. Seidman's et al. (1957) finding seems to indicate alleviation of tension as a function of a shared situation, which, applied to the present study, would suggest lowered tension in the (SwS) group.

In the light of the above discussion, it seems that the experimental design was not well suited to test the second hypothesis. Furthermore, the stronger the first hypothesis was supported, the less possibility there would be of supporting the second.

Correlation of Anxiety Measures

Although the three measures of tension tended to support the first hypothesis, the intercorrelations among the measures were low (Table 7). There is a substantial body of data in the literature which indicates that people differ one from the other with regard to their autonomic response pattern (Gellhorn, 1953). The present finding, suggesting that people, in general, respond to anxiety in many ways, while the individual person may use only one or few characteristic anxiety responses, is in accord with the literature.

Lack of correlation among the physiological measures may also have been a reflection of possible difference in effectiveness of the methods of recording the physiological reactions. For example, spot checking of the blood pressure, as done in this study, may not have guaranteed that the readings represented what the blood pressure was during the work period. The finger perspiration, by contrast, is (within time limits) more of a cumulative and persistent indicator of autonomic change.

Two studies, which related social background to specific pattern of cardiovascular reaction to emotional stress, would demonstrate that characteristic autonomic functions can be learned by an individual (King & Funkenstein, 1957; King & Henry, 1955). However, even if learning of physiological patterns of responses seems possible, the conscious change of such responses, once learned, is not likely to take place. While physiological measures of tensions are thus less under conscious control, psychological measures, such as the employed rating scales, are much

more easily influenced by intentional variations. Lack of correlation between the physiological and psychological measures of anxiety may thus be a function of changes in such factors as apprehension, unfamiliarity, etc., affecting the different measuring instruments differently.

Another possibility of accounting for the lack of correlation of anxiety measures, or at least the absence of a significant correlation between physiological and psychological measures, is in terms of anxious persons' impaired perception of their own anxiety reactions. Verbal reports of anxiety level would thus not correspond with physiological measurements. Support for this suggestion may be found in Table 7. The NS(wNS) group, where least anxiety was present, yielded two of the three significant correlations between the anxiety measures.

SUMMARY

This study was designed to produce an interpersonal situation where communication of anxiety as well as relief of anxiety (comfort) could be demonstrated. It was hypothesized that (a) a person who interacts with a more anxious person will himself become more anxious, and that (b) a person who interacts with a less anxious person will himself become less anxious.

Ss, who were male college students, were randomly paired and assigned to an experimental session or to either of two control sessions. All of the pairs were asked to perform a cooperative work task, a puzzle, the purpose of which was to provide an appropriate means of bringing the partners together in interpersonal interaction. In the experimental session, one S received electric shock while his partner received no shock. The purpose of the shock was to introduce anxiety. The extent to which this anxiety was communicated was measured in three ways: changes in blood pressure, finger sweat, and scores on a rating scale.

The first hypothesis was tested by comparing the anxiety change of the nonshocked Ss in the experimental session with anxiety change of nonshocked Ss in a control session where no S was shocked. The data supported the hypothesis for all three measures of anx-

ity. The second hypothesis was tested by comparing the shocked Ss of the experimental group with shocked Ss in a control group where both members received shock. The findings here did not support the hypothesis for any anxiety measure used.

It appears that the experimental design had less power to show communication of comfort than communication of anxiety. This lack of power could have been due to several considerations: (a) the cues of comfort, that were to be communicated to the anxious S, were not strong enough; (b) the level of anxiety in the control Ss, against which the communicated comfort was compared, was not high enough; (c) the shocked Ss may not have attended to the cues of comfort displayed by their partners; and (d) feelings of comfort cannot be communicated as easily as feelings of anxiety.

An additional finding was that the measures of anxiety did not correlate with each other for the groups of anxious Ss. Where minimal anxiety was present, i.e., in the control group where no shock was given, significant correlations were, however, evident.

REFERENCES

- BENDIG, A. W. The development of a short form of the Manifest Anxiety Scale. *J. consult. Psychol.*, 1956, 20, 384.
- DAVITZ, J. R., & MASON, D. J. Socially facilitated reduction of a fear response in rats. *J. comp. physiol. Psychol.*, 1955, 48, 149-151.
- ERIKSEN, C. W., & WECHSLER, H. Some effects of experimentally induced anxiety upon discrimination behavior. *J. abnorm. soc. Psychol.*, 1955, 51, 458-463.
- FROMM, E. *Escape from freedom*. New York: Farrar & Rinehart, 1941.
- GELLHORN, E. *Physiological foundations of neurology and psychiatry*. Minneapolis: Univer. Minnesota Press, 1953.
- HORNEY, KAREN. *The neurotic personality of our time*. New York: Norton, 1937.
- KING, S. H., & FUNKENSTEIN, D. H. Religious practice and cardiovascular reactions during stress. *J. abnorm. soc. Psychol.*, 1957, 55, 135-137.
- KING, S. H., & HENRY, A. F. Aggression and cardiovascular reactions related to parental control over behavior. *J. abnorm. soc. Psychol.*, 1955, 50, 206-210.
- LIDDELL, H. Some specific factors that modify tolerance for environmental stress. *Res. Publ. Ass. Nerv. Ment. Dis.*, 1950, 29, 155-171.
- LIRZMAN, S. Social activity in anxiety extinction. Unpublished doctoral dissertation, Teachers College, Columbia University, 1955.
- MASSERMAN, J. H. *Behavior and neurosis*. Chicago: Univer. Chicago Press, 1943.
- MOWRER, O. H., LIGHT, B. H., LURIA, ZELLA, & ZELENY, MARJORIE P. Tension changes during psychotherapy, with special reference to resistance. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald, 1953.
- REISER, M. F., REEVES, R. B., & ARMINGTON, J. Effects of variations in laboratory procedure and experimenter upon the ballistocardiogram, blood pressure, and heart rate in healthy young men. *Psychosom. Med.*, 1955, 17, 185-199.
- RUESCH, J., & PRESTWOOD, A. R. Anxiety, its initiation, communication and interpersonal management. *Arch. Neurol. Psychiat.*, 1949, 62, 527-550.
- SEIDMAN, D., BENSON, S. B., MILLER, I., & MEELAND, T. Influence of a partner on tolerance for a self-administered electric shock. *J. abnorm. soc. Psychol.*, 1957, 54, 210-212.
- SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington, D. C.: William Alanson White Psychiatric Foundation, 1947.
- SULLIVAN, H. S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.

(Received November 20, 1959)

PSYCHOLOGICAL TEST COVARIATES OF CONCEPTUAL DEFICIT IN SCHIZOPHRENIA

WILLIAM W. LOTHROP¹

Veterans Administration Hospital, Northampton, Massachusetts

The conceptual ability of the schizophrenic has been the object of a great deal of clinical interest and research in recent years. Many studies, using a number of different techniques, have demonstrated that schizophrenics usually perform less well than normals or neurotics on conceptual tasks. As yet, however, there have been very few systematic attempts to investigate factors which might be related to this deficit. Different studies have described some of these factors, such as inappropriate and/or flattened affect, disturbed intellectual functioning, poor prognosis, chronicity, and hebephrenic diagnosis (Kasanin & Hanfmann, 1938; Meadow & Funkenstein, 1952; Scherer, 1951; Stotsky & Lawrence, 1955), but replication of these findings is generally lacking, and many other factors must remain to be described. The present study represents a preliminary attempt to investigate some of these factors, with the specific purpose of determining whether conceptual deficit in schizophrenia covaries consistently with performance on psychological tests and with certain clinical variables, in the hope of generating hypotheses for future research.

METHOD

The subjects (Ss) used in this study were all male veterans from the VA Hospital in Northampton, Massachusetts, who had been discharged from the hospital with the final staff diagnosis of schizophrenic reaction. Cases with known cortical or neurological

¹ The author wishes to express his gratitude to Isidore W. Scherer, Arnold Trehub, and Henry Openheim of the hospital Psychology Service for their critical evaluations of this manuscript, and to Joseph M. Cuddy and David J. Moriarity, Clinical Psychology trainees, for their help in analyzing part of the data.

The author is now at the University of New Hampshire.

involvement were excluded. A total of 64 patients was selected alphabetically from the Clinical Psychology files who had been given the following tests: Wechsler-Bellevue I, Rorschach, Object Sorting (active part), Sacks Sentence Completion (Form A), Word Association, Bender-Gestalt, Human Figure Drawings, Babcock Memory Paragraph A, Benton Visual Retention Test (1946 form), the Associate Learning test from the Wechsler Memory Scale, and the Paired-Associates from the Hunt-Minnesota Test for Organic Brain Damage. The scoring of all tests was rechecked by the experimenter (E). All patients had been tested within 6 weeks of their admission, and approximately 90% were first admissions to this hospital. Thus the chronic, deteriorated type of schizophrenic frequently associated with conceptual deficit (Benjamin, 1944; Rabin, King, & Ehrmann, 1955; Rapaport, Gill, & Schafer, 1945) is generally missing from this sample, and the results must be interpreted accordingly.

The active part of Rapaport's Object Sorting Test (1945), with slight modifications in administration, was used as the measure of conceptual performance. To obtain one group of Ss showing relatively little conceptual deficit and another group with marked deficit, an elaborate set of criteria was initially developed. It was then found that the same grouping of Ss could be obtained using only the total number of concrete responses and failures. For this report, therefore, the criterion for determining amount of conceptual deficit was the number of concrete responses and failures. The mean number of such responses in this sample was 3.42, with a standard deviation of 1.93. To enhance the magnitude of possible covariant effects on the other variables to be tested, the two extremes of the distribution were selected for analysis. Two or fewer concrete responses were taken to indicate little or no conceptual deficit, based on Rapaport's patrol group data. Twenty-three (36%) of the patients in this sample gave two or fewer such responses. Severe conceptual deficit was arbitrarily designated as five or more concrete responses out of the seven items. Eighteen Ss (28%) fell in this group. Two groups of 18 were obtained by taking the first 18 unimpaired Ss alphabetically and all of the severely impaired Ss. The other variables were then tested against these two groups to determine if significant discrimination could be obtained.

RESULTS

Clinical Variables

In this sample severe conceptual deficit was not related to age at admission; number of school grades completed; occupational level; religion; diagnostic category; whether the illness was rated mild, moderate, or severe by the staff; or whether it was classified as chronic or acute. Three findings significant between the .10 and .05 levels of confidence by Fisher's exact test may deserve replication, however: the severely impaired Ss were more often married, had more often been committed to the hospital, and had spent a greater amount of time in this hospital over all admissions.

Psychological Tests

No relationship could be demonstrated between severe conceptual deficit and Verbal, Performance, or Full Scale IQ on the Wechsler-Bellevue I. Each subtest was also analyzed by the chi square method, but no test reached statistical significance, nor was there any significant difference in intertest scatter between the two groups.

On the Associate Learning test from the Wechsler Memory Scale the unimpaired Ss learned a significantly greater number of associations, $p = .025$ by the exact test. This difference appeared to be primarily a function of the hard associations, there being no difference between the groups in the number of easy associations learned. Results in the same direction were obtained with the hard pairs from the Hunt-Minnesota test, but these results did not attain statistical significance.

None of the three memory tests used (Digit Span, Babcock Memory Paragraph A, and the Benton test) discriminated between the two groups at significant levels of confidence, although in each case the impaired group received lower scores. Actually, both groups showed a deficit on these tests relative to their IQs, reducing the discriminatory powers of the tests.

For the analysis of the Bender-Gestalt test, Ss from the two groups were matched roughly on IQ. Their tests were then presented in pairs to two independent judges, advanced graduate students in psychology, who were

asked to decide which of each pair was the most deviant, without knowing the identity of the records. The test did not discriminate between the two groups. The same procedure was followed with the human figure drawings, again without significant results. Similarly, neither the sex of the first-drawn figure nor the size relation between male and female figures discriminated between the two groups.

On the Word Association Test, a modification of Rapaport's list of 60 words, there were no significant differences between the two groups in number of popular responses or in the number of associations correctly recalled during immediate readministration of the list, regardless of whether the words were "traumatic" or "nontraumatic."

The unimpaired group gave a significantly greater number of responses to the Rorschach, $p = .02$ by the exact test. In view of the possible qualitative and quantitative differences which this might produce in other scoring categories (Cronbach, 1949) a matched-pair technique was then used. Eleven pairs of Ss were obtained by matching to within plus or minus one response, and each of the remaining scoring categories was analyzed by the Wilcoxon matched-pair rank test. Three of the 17 variables tested yielded significance at the .05 level of confidence: M , FC' , and FC , with the unimpaired group giving the greater number of responses in each instance. The unimpaired group also gave a preponderance of M over ΣC which was significant at the .05 level. In addition, there was a tendency for the unimpaired group to give more human and human detail responses ($p = .10$), and fewer animal or animal detail responses ($p = .12$). The unimpaired Ss also tended to receive lower scores on the Buhler-Lefever Diagnostic Sign List (Buhler, Buhler, & Lefever, 1948), indicating less pathology ($p = .10$). Lastly, genetic level scores were computed for each S (Becker, 1956). In 8 of the 11 pairs the unimpaired S received the higher score, indicating a higher level of perceptual development, but one of the reversals was quite large and significance could not be obtained by the Wilcoxon test.

The Sentence Completion test failed to discriminate between the two groups in the number of words used to complete the first 20

items, or in the number of items left unanswered. The content of the individual responses was naturally quite diverse, except for two items. In completing the phrase "To me the future looks . . .," the impaired Ss with but one exception described the future as "good" or "bright," whereas 10 of the unimpaired group described the future as "black," "hopeless," "dull," etc. ($p = .01$ by the exact test). On the item "Compared with others I . . .," 13 of 16 in the unimpaired group could be classified as considering themselves different from other people, usually in a negative sense, while 7 of 11 in the impaired group considered themselves no different from other people ($p = .05$). The responses to the remaining items were too varied to yield significant consistency. It was subjectively felt, however, that the unimpaired group was more frank in describing environmental difficulties. On the four items having to do with mother, for example, the unimpaired Ss made twice as many negative or critical statements than the impaired Ss. The same also held true for statements about the father. Lastly, it was felt that the impaired group tended to deny illness, in that they described fewer fears, fewer interpersonal difficulties, etc. In the absence of a more objective analysis, however, these impressions must be regarded as speculative.

DISCUSSION

The results of this study suggest, first of all, that abnormal concreteness may not be characteristic of all schizophrenics, at least in terms of their Object Sorting performance. While it is true that a majority of this sample gave more concrete responses than would be expected in the normal population, one-third demonstrated little or no conceptual deficit by Rapaport's norms. This finding is consistent with the results of other studies using different measures of conceptual ability (Hanfmann & Kasanin, 1942; Wegrocki, 1940), and re-emphasizes the need to consider individual differences along this dimension.

As regards the relationship of conceptual deficit to performance on psychological tests, the results suggest that a basic intellectual deficit exists which Goldstein (1959) has

termed a "restriction in the use of the highest mental capacity" (p. 147). In this sample, the effects of conceptual deficit became most apparent in unstructured situations, such as the Rorschach, and in tasks involving new learning. In both these cases, as with the Object Sorting Test itself, active and creative thinking is required. Under these conditions the "concrete" schizophrenics demonstrated greater intellectual impoverishment and constriction. This deficit was not significantly manifest on tests involving memory and concentration (both groups showed impairment), perceptual-motor tasks, or on verbal tasks involving relatively stereotyped or familiar material.

The present data do not permit much speculation about the dynamics of conceptual deficit. The impaired group in general seems more deviant, and in some respects performs like an organic group. An alternative interpretation, however, is that conceptual deficit may be related to "defensiveness," as Goldstein (1959) and others have suggested, manifested by a constriction of interpersonal and affective responses as well as intellectual constriction, guardedness, denial, and/or lack of insight. The concrete schizophrenics tended to portray themselves as no different from other people, were allegedly optimistic about the future, and described little stress in their environments on the Sentence Completion test. By contrast with the unimpaired Ss, the statements of the concrete schizophrenics seemed patently unrealistic. Perhaps this denial of illness or lack of insight, if such be the case, may relate to the tendency for these patients to have been more often committed, since the committed patient is often brought to the hospital against his wishes because he sees no reason to be hospitalized. In order to make these speculations tenable, however, further research needs to be done regarding the role of personality factors in conceptual deficit.

SUMMARY

This study attempted to investigate some of the clinical and test covariates of conceptual deficit in schizophrenia, using number of concrete responses and failures on the active part of the Object Sorting Test as the measure of

deficit. Using the test records of patients who had been discharged with the staff diagnosis of schizophrenic reaction, two groups of 18 Ss each were selected, one showing marked conceptual deficit and one which demonstrated little or no deficit. Clinically, no significant differences were found between the two groups in education, occupational level, religion, age at admission, or diagnostic category. The patients with severe deficit, however, tended more often to be married, had more often been committed, and tended to have spent more time in the hospital overall. When compared as to performance on psychological tests, the results suggested that conceptual deficit may be associated with a basic intellectual impoverishment in unstructured situations and in situations requiring new learning, as might have been predicted. In addition, however, it was hypothesized that conceptual deficit may be associated with a defensiveness that is manifested by restricted interpersonal contacts, affective inhibition, and a tendency towards denial of illness and/or lack of insight. These results were thought to indicate that further research into the role of personality factors in conceptual deficit might prove fruitful.

REFERENCES

- BECKER, W. C. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, 53, 229-236.
- BENJAMIN, J. D. A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In J. S. Kasanin (Ed.), *Language and*

- thought in schizophrenia*. Berkeley: Univer. California Press, 1944. Pp. 65-90.
- BUHLER, C., BUHLER, K., & LEFEVER, D. W. Development of the basic Rorschach score with manual of directions. *Rorschach standardization Stud.*, 1948, No. 1.
- CRONBACH, L. J. Statistical methods applied to Rorschach scores: A review. *Psychol. Bull.*, 1949, 46, 393-429.
- GOLDSTEIN, K. Concerning the concreteness in schizophrenia. *J. abnorm. soc. Psychol.*, 1959, 59, 146-148.
- HANFMANN, EUGENIA, & KASANIN, J. S. Conceptual thinking in schizophrenia. *Nerv. ment. dis. Monogr.*, 1942, No. 67.
- KASANIN, J. S., & HANFMANN, EUGENIA. An experimental study of concept formation in schizophrenia. *Amer. J. Psychiat.*, 1938, 95, 35-48.
- MEADOW, A., & FUNKENSTEIN, D. H. The relationship of abstract thinking to the autonomic nervous system in schizophrenia. In P. H. Hoch & J. Zubin (Eds.), *Relation of psychological tests to psychiatry*. New York: Grune & Stratton, 1952. Pp. 131-149.
- RABIN, A. I., KING, G. F., & EHRLMANN, J. C. Vocabulary performances of short-term and long-term schizophrenics. *J. abnorm. soc. Psychol.*, 1955, 50, 255-258.
- RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. I. Chicago: Year Book Publishers, 1945.
- SCHERER, I. W. Prognoses and psychological scores in electroconvulsive therapy, psychosurgery, and spontaneous remission. *Amer. J. Psychiat.*, 1951, 107, 926-931.
- STOTSKY, B. A., & LAWRENCE, J. F. Various Rorschach indices as discriminators of marked and little conceptual disorganizations among schizophrenics. *J. consult. Psychol.*, 1955, 19, 189-193.
- WEGROCKI, H. Generalizing ability in schizophrenia. *Arch. Psychol.*, 1940, No. 254.

(Received November 23, 1959)

THERAPISTS' JUDGMENTS OF APPROPRIATENESS OF PSYCHOTHERAPY FREQUENCY SCHEDULES

DOUGLAS M. MCNAIR AND MAURICE LORR

Veterans Benefits Office, Veterans Administration, Washington, D. C.

At some point early in intensive psychotherapy, therapist and patient, collaboratively or otherwise, decide how frequently treatment sessions will be scheduled. The decision presumably relates crucially to treatment outcome—if one can judge from the clashes within and between different theoretical schools over the matter of optimum frequency. Optimum frequency recommendations range from the orthodox analysts' four or five sessions weekly down to the once weekly appointment schedule commonly employed by nonanalysts. A large proportion of psychotherapy patients are seen even less than once weekly (Hollingshead & Redlich, 1958), although such schedules are rarely recommended or defended in the literature. Decisions about treatment frequency are also presumably based on patient characteristics important to the treatment process—if one accepts the practical and theoretical articles citing patient variables to be considered in setting the treatment schedule (e.g., Wolberg, 1954, pp. 258-259). Certainly, much time is consumed in staff conferences and in discussions between patients and therapists reaching decisions about interview frequency.

The purpose of this study was to examine the following questions concerning therapists' judgments about what frequency schedules are appropriate for their patients:

1. Do patients respond more favorably to psychotherapy if treated on frequency schedules their therapists consider suitable than if treated on schedules their therapists consider unsuitable for their conditions?

2. What patient variables relate to therapist judgments that treatment schedules are suitable or unsuitable for their patients?

3. What other therapist judgments of the patient and the treatment procedure relate to

therapist opinions about suitability or non-suitability of treatment schedules?

METHOD

The present data were collected as part of a larger study to evaluate effects of individual psychotherapy (Veterans Administration Cooperative Study to Predict Response to Individual Psychotherapy). Seven mental hygiene clinics participated in the study.¹ The design of the cooperative study called for random assignment of patients to three interview schedules—twice weekly, once weekly, and biweekly. The vast majority of VA psychotherapy patients are seen on one of these three schedules. One cooperating clinic randomly assigned patients to all three treatment frequencies. The other six clinics employed two of the three schedules—once weekly plus either twice weekly or biweekly. After the first therapy hour, therapists judged whether the assigned schedules were suitable, too frequent, or too infrequent for their patients' conditions. These suitability ratings are the focus of this report.

Subjects. The study included all 133 outpatients who were assigned to treatment and remained for at least 4 months. The average patient was about 35 years old and had attended college. All patients satisfied these criteria for inclusion in the project: male, under age 50, without evidence of brain injury, less than 12 therapy hours in the preceding 3 months, acceptable for intensive psychotherapy. "Intensive psychotherapy" was defined as interviews lasting at least 45 minutes, aimed at changing personal adjustment patterns, and excluding supportive or maintenance therapy. Forty-five patients seen twice weekly, 62 patients seen once weekly, and 26 patients seen biweekly completed the study.

Procedure. Clinics followed usual procedures for prepsychotherapy processing, adding research procedures at each stage. Initial assessment included a research test battery, evaluation by a social worker, and appraisal by the therapist. Reassessment of patients took place after 4 months of psychotherapy. Each clinic staff made the assignments of patients to

¹ We are indebted to the staffs of the Veterans Administration Mental Hygiene Clinics in the following cities for cooperating in the collection of data for this study: Albany, Boston, Bridgeport, Buffalo, Chicago, Denver, and Hartford.

therapists, and interview frequencies were determined randomly after selection of the therapists. In all instances that the assigned interview schedule had to be altered, the patient was dropped from the project. Psychologists were therapists for 45% of the patient sample, psychiatrists for 23%, and social workers for 32%.

Measures. The patient, therapist, and social worker measures cited in this report are briefly described below.

Patient measures:

Anxiety Scale—a slightly modified form of Taylor's (1953) MAS

Ego Strength—56 items from Barron's (1953) 66-item scale from the MMPI (10 items omitted due to overlap with other measures or because they were potentially offensive questions about religious beliefs)

Symptom Checklist—an inventory of 20 common somatic complaints

Self-Rating—15 five-point graphic scales designed to measure self-satisfaction

K scale—15 items from the MMPI K scale

Sociability—a 15-item questionnaire from the Guildford-Zimmerman Temperament Survey

Behavior Disturbance—50 questionnaire items adapted from an inventory by Applezweig, Dibner, and Osborne (1958) to measure antisocial behavior and psychopathic tendencies

Word Fluency—based on a test to measure the Word Fluency Factor in Thurstone's Primary Mental Abilities Test

Education—nine-point scale of level of schooling completed

Year of Birth

Therapist ratings or reports:

Change Report-IR—inventory of 24 specific changes in the patient-therapist interview relationship, each item describing a change commonly reported by therapists as accompanying psychotherapy

Change Report-PG + SR—inventory of 68 specific positive gains and symptom reductions, each item describing a change commonly reported by therapists as accompanying psychotherapy

Severity Rating—rating on seven four-point graphic scales related to severity of illness (Lorr, Holsopple, & Turk, 1956)

Global Severity—four-point graphic scale for rating over-all severity of disorder

Liking for Patient—four-point graphic scale of degree of liking for the patient as a person

Interest in Problem—four-point graphic scale of degree of interest in the problem presented by the patient

Motivation for Treatment—four-point graphic scale of patient's degree of motivation for treatment

ICL—each of four measures describing the patient on a set of 32 adjectives from an interpersonal checklist developed by La Forge and Suczek (1955): Dominance, Hostility, Submissiveness, Affection.

Social worker rating:

Severity Rating—same scales as the therapist measure

RESULTS

Suitability and Outcome

To determine if patients on schedules judged suitable respond more favorably to psychotherapy than patients on schedules judged unsuitable, the total sample was dichotomized into Groups S and U. Group S consisted of 97 patients on schedules the therapists considered suitable. Group U included 36 patients on schedules the therapists considered unsuitable—either too frequent or not frequent enough. All treatment frequencies were pooled for the analysis, since, for these patients, actual assigned treatment frequency was unrelated (as shown by statistical tests) to outcome.

Results of analyses of covariance for 10 criteria are summarized in Table 1. Pretreatment scores on the criteria were controlled in all analyses except the two Therapist Change Reports—where there were no initial scores. One-way analysis of variance results are presented for these two measures. One-tailed tests of significance were made, since greater improvement was hypothesized for Group S. Column 2 of Table 1 indicates the hypothesized directions of the mean differences.

The results indicated one significant change in the predicted direction on a patient measure (*K* scale), one reversal (Self-Rating), and one positive predicted result on a therapist measure (Change Report-IR). The *K* scale difference suggests that Group S patients were more ready to report problems after 4 months of psychotherapy. Even though statistically significant, the mean difference was so slight as to be of doubtful clinical importance. Therapists reported significantly more changes in the patient-therapist interview relationship for Group S. The Change Report-IR included such items as "more able to report dreams and embarrassing material," "relies less on therapist for reassurance," and "misses fewer therapy appointments."

Subsequently, Groups S and U were pooled and *t* tests were made of correlated pre-post-treatment differences on the criteria (except Change Index-PG + SR) which were not significant by the above analyses. These tests were made to determine if the entire group showed other changes over the course of treat-

TABLE 1
FINAL ADJUSTED MEANS AND *F* TESTS FOR GROUP S AND GROUP U
(Initial scores controlled)

Criterion	Hypothesis	Group S		Group U		<i>F</i>
		<i>N</i>	Mean	<i>N</i>	Mean	
Patient measure:						
Anxiety	<i>S</i> < <i>U</i>	97	29.32	36	27.87	1.462
Ego Strength	<i>S</i> > <i>U</i>	97	32.08	36	32.25	0.030
Symptom Checklist	<i>S</i> < <i>U</i>	97	7.78	36	7.33	0.639
Self-Rating	<i>S</i> > <i>U</i>	96	46.09	36	48.51	5.540 ^a
<i>K</i> Scale	<i>S</i> < <i>U</i>	97	5.30	36	6.09	2.812*
Sociability	<i>S</i> > <i>U</i>	97	7.43	36	7.82	0.837
Therapist report:						
Change Index-IR	<i>S</i> > <i>U</i>	97	9.58 ^b	34	7.03 ^b	7.244**
Change Index-PG + SR	<i>S</i> > <i>U</i>	97	12.39 ^b	34	10.68 ^b	1.047
Severity Rating	<i>S</i> < <i>U</i>	97	2.09	34	2.04	0.519
Social worker report:						
Severity Rating	<i>S</i> < <i>U</i>	97	1.96	34	2.02	0.486

^a Reversal, not significant by one-tailed test; $p < .05$ if two-tailed test had been specified.
^b Means are actual, not adjusted. There were no initial scores on this variable.

* Significant at .05 level.

** Significant at .01 level.

ment which were not associated with the suitability of the treatment schedule. These data are not presented in detail, but none of the four patient measures thus tested proved significant. The Therapist Severity Rating showed a significant decrease ($p < .01$) as did the Social Worker Severity Rating ($p < .05$).

Following the above analyses, additional sets of comparisons of response to treatment were made. The 11 patients rated as seen too frequently on the twice weekly schedule were compared with the 34 patients rated suitable for twice weekly interviews; and the 14 patients rated as seen too infrequently on biweekly schedules were compared with the 12 patients rated suitable for biweekly interviews. (A similar comparison could not be made for the once weekly group as too few patients were rated as scheduled either too frequently or too infrequently.) Thus all patients in each comparison were rated in relation to the same standard—actual treatment frequency. It was thought that differences might have been obscured in the main analyses by pooling all unsuitable cases into one

group regardless of actual treatment frequency and regardless of whether they were considered seen too often or too infrequently. As in the main analysis, greater improvement on each frequency schedule was hypothesized for patients on suitable schedules.

Covariance analyses on the same 10 criteria listed in Table 1 were run for the twice weekly and biweekly groups. These data are not presented in detail, but there was one significant difference beyond the .05 level in each group—*K* scale in the twice weekly group and Change Report-IR in the biweekly group. There were no consistent trends of results on the other criteria. The absence of trends and the finding of only two significant results in 20 comparisons make it appear doubtful that suitability ratings predict response to psychotherapy over this period of treatment.

A natural question might arise as to whether a disproportionate number of patients on unsuitable schedules had terminated treatment before the 4-month re-evaluation date. A chi square test indicated that the 133 and 175 patients who, respectively, completed and terminated the study did not differ signifi-

TABLE 2

INITIAL MEANS AND *F* TESTS FOR THERAPISTS' RATINGS OF GROUP S AND GROUP U
ON TWICE WEEKLY AND BIWEEKLY TREATMENT SCHEDULES

Variable	Twice Weekly Interviews			Biweekly Interviews		
	Means		<i>F</i>	Means		<i>F</i>
	Group S (<i>N</i> =34)	Group U (<i>N</i> =11)		Group S (<i>N</i> =12)	Group U (<i>N</i> =14)	
Liking for Patient	2.50	1.82	8.250**	2.25	2.21	0.016
Interest in Problem	2.65	2.27	2.282	2.33	2.43	0.154
Motivation for Treatment	2.68	2.64	0.029	2.33	2.57	0.624
ICL-Dominance	3.26	3.55	0.057	3.42	2.93	0.108
ICL-Hostility	6.64	7.03	0.056	6.17	6.14	0.000
ICL-Submissiveness	6.24	10.09	5.055*	7.67	5.50	1.134
ICL-Affection	2.76	3.55	0.566	4.00	2.64	0.574
Global Severity	2.88	3.36	5.877*	3.00	2.79	0.576

* $p < .05$.** $p < .01$.

cantly in the number of patients on unsuitable schedules.

Relation of Suitability Ratings to Patient Characteristics

Initial characteristics of patients on suitable and unsuitable schedules were compared. *F* tests were made to determine if Groups S and U on either the twice weekly or biweekly schedules differed initially on the 10 patient measures included in the study. Most of these characteristics have been cited in the literature as related to patient response to psychotherapy—e.g., best prospects for psychotherapy are often described as relatively fluent, young, educated, nonpsychopathic, anxious, etc. There were, however, no significant differences in patient characteristics between Groups S and U on the twice weekly schedule—where Group U was described as needing less frequent interviews. Nor were there differences between Groups S and U on the biweekly schedule—where Group U was described as needing more frequent sessions.

Suitability Ratings and Other Therapist Ratings of the Patient

Suitable and unsuitable patient groups were also compared with respect to their therapists' judgments about them, and the results are summarized in Table 2. All therapist ratings

and reports included in Table 2 were made after the initial therapy hour. In the twice weekly group, patients rated unsuitable—as needing less frequent interviews—were liked significantly less by their therapists. Therapists described Group U as less likable than most patients, and Group S as more likable than most patients. The twice weekly Group U was also rated as significantly more severely ill than Group S. The ICL-Submissiveness difference indicates therapists described Group U patients significantly more often with adjectives such as “self-punishing,” “passive,” “shy,” and “dependent.”

Therapists' ratings and reports did not differentiate suitable and unsuitable groups on the biweekly schedule. Table 3 may help ex-

TABLE 3
NUMBER OF PATIENTS RATED SUITABLE
AND UNSUITABLE FOR THREE
TREATMENT SCHEDULES

Therapists' Rating	Assigned Treatment Schedule			Total
	2×wk.	1×wk.	Biwk.	
Suitable	34	51	12	97
Unsuitable	11	11	14	36
Total	45	62	26	133

Note.— $\chi^2 = 12.33$, $df = 2$, $p < .01$.

plain the rather consistent absence of differences between Groups S and U on the biweekly schedule. In Table 3 the relation of suitability ratings to actual assigned treatment frequency is presented. Chi square is significant at $p < .01$, but the frequency of unsuitable ratings of patients on the biweekly schedule produces the significant chi square. Thus, an unsuitable rating for a patient in the biweekly group appears to reflect therapist dissatisfaction with a treatment schedule providing for fewer interviews than the conventional once-a-week practice. It reflects little else as far as can be determined here.

DISCUSSION

There is little evidence that patient response to psychotherapy differs depending upon whether the treatment frequency is considered suitable or unsuitable. A report by the therapist on the number of changes observed in the interview relationship is the single indicator of greater change by the suitable group. It is questionable that the number of interview changes reflects a true patient difference in treatment response. Interview relationship patterns are modified by therapists as well as patients, and therapist satisfaction or dissatisfaction with a treatment schedule could be a factor influencing this pattern. It thus appears doubtful that the therapist can tailor the interview schedule to fit the individual psychotherapy patient—if the criterion for “fit” is favorable treatment response. Of course, certain obvious qualifications—the sample, the length of the treatment period studied, the fact that therapists judged suitability early in treatment—apply.

A pertinent question is: were therapists unable to judge the appropriateness of treatment schedules because therapy had no effect on this sample, or were they simply unable to judge accurately or validly how many interviews per week were necessary to obtain that effect? The results do not permit a decisive answer. One reason is that there is lack of confirmation between patient measures and observer reports concerning whether the sample as a whole showed significant changes from initial status over the treatment period. Independent ratings by therapists and social

workers indicated significantly decreased severity of patients' disorders as compared with pretreatment status. Yet none of the four directly assessed patient criteria showed any such significant changes. There are at least two plausible explanations for the lack of confirmation: first, patients and observers may be reporting correctly but responding to different events; second, there may be a systematic bias on the part of the observer clinicians to report improvement after several months of therapy. Because of the possibility of rater bias, there is some doubt as to whether this sample of patients changed appreciably from initial status. At any rate, it is clear that the two severity ratings did not change as a function of appropriate or inappropriate treatment schedules even though they did change significantly from initial status.

Alternatively, therapy on “suitable” and “unsuitable” schedules represents psychotherapy conducted under two conditions: effective and ineffective. Thus, an equally pertinent question is: do patients treated under conditions therapists regard as effective actually respond, or respond more favorably than patients treated under less effective conditions? It is this question that therapists often raise in interpreting studies where patients are randomly assigned to different treatments. The evidence suggests that over the period studied, both conditions are, perhaps, slightly effective, but the supposedly favorable condition has no advantage.

Similarly, the therapists' opinions about suitability of an assigned treatment schedule failed to relate to patient characteristics assessed by administering objective psychological tests or obtaining actuarial information. Ten such characteristics were included, and most of them have been cited in the literature as related to readiness, prognosis, or ability to participate in the task of psychotherapy.

In a previous study, Hollingshead and Redlich (1958, pp. 269–270) found social class related to scheduled frequency of psychotherapy interviews. In an unpublished survey, Lorr and McNair found that education, an important component of social class indices, related to scheduled interview fre-

quency. The present findings might appear to partially disconfirm the above findings, since education did not relate to therapists' preferences for more or less frequent interviews with the patient. The apparent disagreement may be due to the fact that the former of the above studies was based on private practitioners' patients and the latter was based on a broad sample of the total VA Mental Hygiene Clinic population, whereas the present study included only patients who met certain screening criteria for inclusion in the study, including the criterion of acceptability for intensive psychotherapy. Secondly, early terminators were eliminated from the present sample which included only patients remaining in treatment for 4 months or longer. Thus, the finding that therapists' suitability judgments did not relate to education or other patient characteristics may hold only for patients who remained in therapy at least 4 months. It can be argued, however, that the eliminated patients are probably not amenable to the usual methods of psychotherapy, and the question of effectiveness of different psychotherapy schedules with such patients is hardly *apropos*.

The concentration of significant differences in therapists' ratings and reports suggests that therapists' reasons for preferring one treatment schedule to another with given patients lie more with the therapists than with the patients. The preferences appear to be associated with personal reactions to patients (such as liking), with presumably objective clinical judgments about patients (such as severity of illness), and with practical and theoretical opinions about the task of doing therapy (such as an aversion to biweekly interview schedules). Whether therapists, in practice, actually arrange or schedule interview frequencies on the basis of such factors is not answered by this study. Therapists in the study had no control over the treatment frequency as long as the patient remained in the study. They simply rated whether an assigned interview schedule of therapy was appropriate for the patients' problems, and indicated preferences for more or less frequent interviews. The importance of therapist factors in actual practice is certainly implied, but no more than that.

An "ideal" or optimum treatment schedule might be described as the one which most effectively shapes patient behavior in the direction of selected criteria improvement. Whether an optimum schedule can be selected for an actual patient is an exceedingly complex question which involves, in something like the following order: knotty and unanswered problems concerning criteria of improvement, questions of whether different interview schedules actually produce different results, questions of whether kinds or types of patients respond differentially to different schedules, and problems of determining predictors of those patients who respond "best" on different schedules. The present results imply, as might naturally be expected, that in the absence of more valid answers, therapists rely on their own learnings and judgments in expressing preferences for interview schedules. The evidence is unimpressive that they can accurately predict which schedules will most effectively alter the patients' behavior, but only future research can provide any sounder bases for making such decisions.

SUMMARY

One hundred and thirty-three outpatients accepted for psychotherapy at seven clinics were randomly assigned to three treatment frequencies. Patients were assessed before treatment and following 4 months of psychotherapy. After the initial interview, therapists rated suitability of the assigned treatment schedule for the patients' conditions. The hypothesis that patients treated under conditions considered favorable by therapists (Group S) would improve more than patients treated under unfavorable conditions (Group U) was tested on 10 criteria. The evidence did not indicate that these groups responded differentially to 4 months of psychotherapy. As evidence of any treatment effect with this sample of patients is indefinite, it remains indeterminate whether therapists can or cannot judge accurately how many interviews per week are needed to obtain optimal treatment effects. Treatment effects must be demonstrated in order to assess the validity of therapists' judgments.

Suitability ratings appear to be associated

with therapists' personal reactions to patients, clinical judgments about patients, and opinions about the task of doing therapy. Suitability ratings did not relate to 10 objectively assessed patient characteristics.

REFERENCES

- APPLEZWEIG, M. H., DIBNER, A. S., & OSBORNE, R. L.
PEAQ: A measure of psychopathic behavior. *J. clin. Psychol.*, 1958, 14, 26-30.
- BARRON, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, 17, 327-333.
- HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.
- LA FORGE, R., & SUCZEK, R. The interpersonal dimension of personality: III. An interpersonal checklist. *J. Pers.*, 1955, 24, 94-112.
- LORR, M., HOLSOPPLE, J. Q., & TURK, ELIZABETH. A measure of severity of illness. *J. clin. Psychol.*, 1956, 12, 384-386.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- WOLBERG, L. R. *The technique of psychotherapy*. New York: Grune & Stratton, 1954.

(Received November 23, 1959)

THE RELATIONSHIP OF FACTORS IN PARENTAL RATINGS OF SELF AND EACH OTHER TO THE BEHAVIOR OF KINDERGARTEN CHILDREN AS RATED BY MOTHERS, FATHERS, AND TEACHERS

WESLEY C. BECKER¹

University of Illinois

As part of a long term investigation of parent-child relationships, a group of psychologists from the University of Illinois Psychological Clinic have been studying a number of approaches to the problems of measuring aspects of parent and child behavior critical in understanding child development. These approaches have included adaptations of interview rating procedures developed at the Fels Institute (Baldwin, Kalhorn, & Breese, 1949) as well as those developed by Sears (Sears, Maccoby, & Levin, 1957), projective tests, the Guilford personality inventories, parent attitude questionnaires, direct observations of parent-child interaction under controlled conditions, objective measures of child behavior, and ratings of child behavior.

The present report discusses the development of two rating instruments, one for child behavior and one for parent behavior, and the relationships between the measures derived from these two instruments.

SUBJECTS

The subjects (Ss) were drawn primarily from five PTA kindergartens in Champaign, Illinois. Out of 123 families represented in these kindergartens, 64 agreed to participate. Teachers' ratings were obtained

on all 123 children, while complete data were obtained on 60 families. Teachers were paid \$2.50 per hour and each family was paid \$10 for participating in the larger study. Eleven families with kindergarten children who came to the Psychological Clinic for help with behavior problems were added to the sample for the analyses of relationships between parent and child measures. None of these additional Ss was a severe problem and they did not differ from the nonclinic Ss on such variables as age, education, and occupational level of father. Table 1 presents some of the sample characteristics for the 71 participating families and for 59 of the nonparticipating families. As indicated in Table 1, the sample is primarily middle class, and education of parents was a selective factor influencing participation. The sex division of the children was almost even in both groups.

TABLE 1
SAMPLE CHARACTERISTICS

Variables	Participants (N = 71)		Nonparticipants (N = 59)	
	M	σ	M	σ
Age of mother	32.4	4.93	32.8	4.92
Highest grade completed, mother	14.2	1.93	12.7	1.72
Age of father	34.7	5.71	35.7	5.82
Highest grade completed, father	15.8	2.46	14.5	3.13
Warner Occupational Level, father (1 = professional, 6 = laborer)	2.25	.97	2.86	1.38
Age of child in months at time of parent ratings	69.3	3.85	69.8	3.71

¹ The study is a part of a cooperative project with Donald R. Peterson, Donald J. Shoemaker, Leo A. Hellmer, and Zella Luria. The writer wishes to express his appreciation to the State of Illinois Department of Public Welfare for its generous financial support, and to Sally Beck, Rogers Elliott, Jacqueline Goldman, Ronald Krug, Irvin Moelis, Geraldine Piorkowski, and Kenneth Stark for their assistance in gathering and analyzing the data.

A COMPARISON OF CHILD PERSONALITY FACTORS OBTAINED FROM MOTHER, FATHER, AND TEACHER RATINGS

This section discusses the development of a rating instrument designed to broadly sample the dimensions of child behavior. In addition, the comparability of mother, father, and teacher ratings in terms of *factor loadings* and in terms of *factor scores* is evaluated. It is quite possible that though mothers, fathers, and teachers (using a common language system) might agree in terms of factor loadings on which rating variables covary, all three groups could markedly disagree on how to apply the ratings to a given child. This latter situation could arise because of different biases, different frames of reference, or because each of the three groups observes different modal sets of behavior. While several previous studies have used both parent and teacher ratings (Cattell & Coan, 1957a, 1957b; Peterson & Cattell, 1958), the writer believes that this paper represents the first systematic effort to compare ratings in terms of both factor loadings and factor scores, and to separately evaluate ratings by mothers and fathers.

The Child Rating Schedule

The child rating schedule consisted of 72 bipolar, seven-point rating scales with antonym pairs of adjectives defining the extremes. Scales were selected to sample the personality domain as outlined by Cattell (1957), as well as the preliminary factors found by Osgood and Becker (unpublished) in the application of the semantic differential method to the study of personality. A number of variables particularly relevant to the behavior of five-year-olds, such as "ease of discipline," were also included. To save space, a formal listing of the variables will not be given. However, their general nature may be ascertained from the tables to follow. Average time required for making the ratings was 5-10 minutes per child.

The Factor Analyses

The general procedure followed was to find (with the aid of a high speed computer) the intercorrelations among the 72 variables, to extract centroid factors, using fixed-unit com-

munalities, and then to rotate to orthogonal simple structure using the varimax criterion (Kaiser, 1958). Separate analyses were determined for the mother ratings ($N = 64$), the father ratings ($N = 60$), and the ratings by each of two sets of teachers ($N = 123$). Thus, four separate factor analyses were involved.

Six teachers participated in the teacher ratings: two rating 60 children, two rating 47 children, and two rating 16 children. The pairs of teachers with the larger groups actually had two classes, one in the morning and one in the afternoon. Before combining teacher rater groups, mean differences in ratings for the 72 variables were examined both within groups and between groups to decide whether it was reasonable to convert to standard scores by teachers before combining. Combining standard scores has the advantage over combining raw scores in that it is then known that the covariance is not due to mean differences in rater frames of reference. However, if there are real mean individual differences between groups for some variables, this variance will be lost. The average absolute differences in mean rating between groups were .36, .42, and .50; while the average absolute differences in mean rating between teachers within groups were .59, .42, and .68. Thus, group differences were found to be less than rater differences within groups. As a result, scores were standardized by teachers before combining into Teacher Rater 1 (T_1) and Teacher Rater 2 (T_2) groups.

In the first stage of the analysis, 10-centroid factors were extracted from each correlation matrix. Generally, the decision of how many factors to rotate was based on an examination of the variance extracted by each factor. The cut was made about where the variance percentage leveled off and seemed to fluctuate randomly. However, an attempt was made to keep one more factor for rotation than was likely to be useful, the weakest factor to be discarded after rotation. This procedure is permissible since varimax does not attempt to distribute variance among "insignificant" factors as do some other analytic methods (Kaiser, 1958). Seven factors were rotated in each case except for fathers, where eight factors were rotated.

The variance contributions of the rotated

factors led to the discarding of one factor from each parent analysis, and one from the T_1 analysis. Two factors were discarded from the T_2 analysis. Examinations of factor plots indicated that two father factors contained but one substantial hyperplane, so a hand rotation was made to eliminate an additional father factor. In a similar way, a small factor was eliminated from the mother analysis and from T_1 . The only other modification made of the analytic solution was the rotation of two factors in the T_2 analysis (Hostile-Withdrawal and Submission) by 24 degrees so that they would better line up with the factors found in the other three solutions. Needless to say this was an arbitrary procedure which could also have been used to improve other matchings. On the other hand, failure to have made this rotation would not appreciably have affected the results reported below.

Matching of Factors

There are many reasons why one might fail to find matchings for *particular* rating scales, e.g., different samples of behavior to observe, different rater concept systems, different emotional biases of raters, different interpretations of the ambiguous terms in the rating scales. Because of these possibilities, the writer did not spend undue time searching for the "best" simple structure nor in computing "refined" statistics on the goodness of factor matching (see Cattell, 1957, for a discussion of this problem). The more practical question amounted to whether or not the number of salient variables showing a reasonable match was sufficient to give a reliable factor measure.

The salient variables for four factors showing reasonable matching in all four analyses are presented in Tables 2 to 5. Since unreliability of ratings would lead to some instability in factor loadings, even if parents and teachers were using common systems, a common salient variable was defined as a loading of .30 or higher appearing in three of the four factors to be compared. Throughout the tables which follow these conventions were used: (a) factor loadings have been made positive and the variable descriptions reversed where necessary; (b) factors have been named according to the direction of a high score; (c)

TABLE 2
COMMON FACTOR 1—HOSTILE-WITHDRAWAL

Variables	Factor Loadings of Salient Variables				
	MI	FI	T_1 II	T_2 IV'	Average
4. Sociable-unsociable	79	76	78	79	78
19. Warm-cold	71	82	81	68	76
8. Happy-depressed	69	77	76	59	70
21. Responsive-alloof	61	77	76	68	70
10. Loving-not loving	64	81	71	55	68
26. Colorful-colorless	49	83	72	56	65
3. Extraverted-introverted	37	60	76	83	64
33. Interesting-boring	53	81	66	53	63
18. Optimistic-pessimistic	60	65	66	41	58
29. Real-unreal	64	73	58	20	54
12. Trusting-distrusting	58	12	47	37	38
23. Soft hearted-hard hearted	38	45	23	34	35
17. Curious-uninquiring	58	86	56	47	(62)
2. Active-inactive	19	85	70	63	(59)
22. Adventurous-timid	27	64	69	68	(57)
50. Interested-bored	44	74	44	47	(52)
72. Adjusted-maladjusted	59	69	37	35	(50)
16. Quick-slow	35	69	54	37	(49)
34. Confident-subjectively inferior	49	45	48	42	(46)
69. Effective-ineffective	39	73	41	23	(44)
43. Likes school-dislikes school	32	53	41	45	(43)
35. Formed-formless	38	62	53	20	(43)
37. Noisy-quiet	00	46	49	57	(38)
27. Outgoing-self-centered	27	30	45	36	(34)

M, F, T_1 , T_2 , have been used to refer to mother, father, Teacher Rater 1 and Teacher Rater 2, respectively; (d) an apostrophe has been added to a factor title to signify that a rotation was made beyond the varimax solution; and (e) as discussed later, parentheses

TABLE 3

COMMON FACTOR 2—RELAXED DISPOSITION

Variables	Factor Loadings of Salient Variables				Average
	MII	FVII'	T ₁ VII	T ₂ III	
54. Tense-relaxed	75	56	69	79	70
64. Nervous-placid	68	35	69	80	63
46. Excitable-calm	57	63	50	67	59
56. Emotional-objective	68	19	46	66	50
49. Anxious-nonchalant	42	43	58	47	48
42. Fluctuating-stable	61	51	22	50	46
40. Fearful-not fearful	34	35	56	42	42

have been placed around factor loadings of variables not actually used in the computation of factor scores.

Common Factor 1, presented in Table 2,

TABLE 4

COMMON FACTOR 3—LACK OF AGGRESSION

Variables	Factor Loadings of Salient Variables				Average
	MV'	FII	T ₁ V'	T ₂ V	
11. Demanding-not demanding	72	60	72	60	66
30. Prone to anger-not prone to anger	70	70	53	47	60
15. Jealous-not jealous	64	33	64	64	56
71. Prone to tantrums-not prone to tantrums	54	60	67	28	52
20. Impatient-patient	46	64	65	28	51
28. Irritable-easy going	63	22	59	57	50
47. Conceited-self-critical	29	58	38	54	45
27. Self-centered-out-going	43	11	56	57	42
60. Exhibitionistic-modest	27	38	51	34	38
5. Cruel-kind	37	22	49	39	37
61. Difficult to discipline-easily disciplined	57	64	45	10	(44)
52. Disobedient-obedient	48	44	38	02	(33)
35. Formed-formless	39	35	20	35	(32)
38. Masculine-feminine	47	31	04	38	(30)

loads most highly on such variables as unsociable, cold, aloof, not loving, introverted, pessimistic, and hard hearted. This factor is very similar to Cattell's (1957) behavior rating Factor A, Cyclothymia—Schizothymia, although it is more general than A and includes most of the elements of Cattell's F factor, Surgency—Desurgency. This factor will be referred to as *Hostile-Withdrawal* versus *Warm-Extraversion*.

Common Factor 2 (Table 3) loads highly on relaxed, placid, calm, objective, stable, nonchalant, and not fearful. This factor contrasts a *Relaxed Disposition* versus a *Nervous Disposition* and closely parallels Cattell's C factor, Ego Strength versus Proneness to Neuroticism.

Common Factor 3 (Table 4) is described by the following variables: not demanding, not prone to anger, not jealous, not prone to tantrums, patient, and easy going. The parallel factor in Cattell's system is Factor D, Excitability, Insecurity versus Emotional Maturity. Common Factor 3 will be referred to as *Lack of Aggression*.

Common Factor 4, as indicated in Table 5, is clearly a *Submission-Dominance* factor. Again there is a parallel to Cattell's Factor E, Dominance.

TABLE 5

COMMON FACTOR 4—SUBMISSION

Variables	Factor Loadings of Salient Variables				Average
	MIII	FVI	T ₁ I	T ₂ I'	
58. Strong willed-weak willed	57	61	47	60	56
59. Independent-dependent	54	83	21	54	53
7. Dominant-submissive	57	50	55	44	52
22. Adventurous-timid	60	47	38	53	50
13. Tough-sensitive	48	-10	68	72	44
37. Noisy-quiet	39	05	54	58	39
2. Active-inactive	56	-01	32	38	31
61. Difficult to discipline-easy to discipline	29	33	73	77	(53)
40. Not fearful-fearful	49	35	30	07	(30)

TABLE 6

FACTOR COMMON ONLY TO TEACHER RATINGS—
SCHOOLROOM INTELLIGENCE

Variables	Factor Loadings of Salient Variables		
	T ₁ III	T ₂ II	Average
9. Dull minded-intelligent	79	85	82
55. Subject to distraction-able to concentrate	80	82	81
44. Poor memory-good memory	76	83	80
69. Ineffective-effective	75	76	76
31. Meaningless-meaningful	75	69	72
16. Slow-quick	56	68	62
34. Subjectively inferior-self- confident	64	61	62
50. Bored-interested	62	62	62
35. Formless-formed	69	39	54
39. Shallow-deep	61	44	52
17. Uninquiring-curious	60	38	49
70. Disorganized-organized	83	81	(80)
63. Irresponsible-responsible	73	82	(78)
66. Infantile-adult-like	67	73	(70)
6. Conscienceless-conscientious	51	71	(61)
59. Dependent-independent	72	43	(58)
65. Not helping-helping	53	56	(54)
29. Unreal-real	48	58	(53)
48. Disorderly-neat	51	53	(52)
72. Maladjusted-adjusted	48	52	(50)
33. Boring-interesting	51	45	(48)
43. Dislikes school-likes school	34	53	(44)
67. Obstructive-cooperative	39	50	(44)
58. Weak willed-strong willed	38	34	(36)

The final area of some common pattern is that of intelligence. An intelligence factor is clearly revealed in both sets of teacher ratings. Table 6 presents the salient variables and indicates not only that the agreement between teachers is exceptionally high, but also that the factor is much broader than what is usually measured by an intelligence test. Other variables load on this factor which might be expected to influence excellence of schoolroom adjustment, namely, responsibility, independence, morality, and adjustment. This factor will be referred to as *Schoolroom Intelligence* to distinguish it from test intelligence.

Since the separate analyses had provided considerable evidence for four common fac-

tors across all raters, an attempt was made to reproduce these factors from the averaged correlation matrix formed by treating each rater's data as additional cases in computing the correlations. Thus, each element of the correlation matrix was based on 370 (123 + 123 + 64 + 60) elements. Because of the differences in *N*, the teacher ratings were doubly weighted in this analysis. Standard scores were used in computing these averaged correlations so that mean rater differences would not effect the correlations. Six varimax factors were found, only five of which were interpreted. The factors referred to as Hostile-Withdrawal, Relaxed Disposition, and Lack of Aggression were clearly extracted. However, the Submission factor was fused with the Schoolroom Intelligence factor. On the other hand, a factor emerged which had previously appeared as a more general factor cutting across the diagonals of the factor plots for Lack of Aggression, Submission, and Relaxed Disposition. The salient loadings on this factor are presented in Table 7. The factor is defined by such variables as disobedient, irresponsible, obstructive, difficult to discipline, disorganized, conscienceless, and lying. This

TABLE 7

COMMON FACTOR 5—CONDUCT PROBLEMS

Variables	Factor Loadings
52. Obedient-disobedient	65
63. Responsible-irresponsible	65
67. Cooperative-obstructive	65
61. Easily disciplined-difficult to discipline	60
70. Organized-disorganized	60
65. Helping-not helping	55
66. Adult-like-infantile	55
48. Neat-disorderly	50
43. Likes school-dislikes school	45
62. Attention avoiding-attention seeking	45
6. Conscientious-conscienceless	40
53. Truthful-lying	30
72. Adjusted-maladjusted	(55)
55. Able to concentrate-subject to distraction	(55)
60. Modest-exhibitionistic	(45)
50. Interested-bored	(41)
69. Effective-ineffective	(40)

factor is very similar to a *Conduct Problem* factor found on these same Ss by Peterson using another rating instrument. Although it would be possible to assess this variable by using a weighted combination of the four common factors above, it was decided that this factor was important enough clinically to merit direct scoring.

Computation of Factor Scores

At this point a decision was made to score for: (a) the four common factors, Hostile-Withdrawal, Relaxed Disposition, Lack of Aggression, and Submission; (b) the non-orthogonal Conduct Problem factor; and (c) Schoolroom Intelligence based only on teacher ratings. The variables used in scoring each factor are indicated in Tables 2 to 7 by the lack of parentheses around the factor loadings. The goal in scoring was to avoid counting any variable for more than one factor. The most obvious reason for doing this was to avoid building in spurious correlations between factors. There is an additional less obvious reason which is related to the first. Even though orthogonal factors are being used (except for Conduct Problem), the factor plots repeatedly indicated some obliqueness in the hyperplanes. Because of this obliqueness, it is to be expected that the factor scores derived from the addition of salients will actually be correlated and will give an estimate of the obliqueness inherent in the structuring.

In computing factor scores, the method of standard score addition of salient variables, weighted according to the size of factor loadings, was used (see Cattell, 1957, for a discussion of various ways of computing factor scores).

Comparison of Factor Scores

The comparison of factor loadings indicated a fair degree of convergence of factor patterns. This convergence does not tell one, however, whether parent and teacher ratings on these factors can be considered equivalent. This question can only be answered by examining the correlations among the factor scores.

Table 8 presents the intercorrelations for the five common factors, plus the Schoolroom

Intelligence factor based on an *N* of 71. Correlations above .23 are significant at the .05 level and those above .31 are significant at the .01 level. In preparing Table 8, the factor scores for the two teachers were averaged. Uncorrected reliabilities of the averaged teacher factor scores are presented in teacher-teacher diagonal.²

Table 8 permits the reader to examine the obliqueness of factors and the comparability of factors across raters. The correlations critical for examination of the comparability of parent and teacher factor scores are italicized in the diagonals of Table 8. The average correlation (using the *z* transformation) for Factors 1 to 5 was .76 between teachers, .52 between parents, .31 between mothers and teachers, and .28 between fathers and teachers. The average correlation for *pooled* parent ratings with the pooled teacher ratings was .34. These findings clearly indicate that parent and teacher ratings have very little in common (8-12% of the variance). However, the reader should note that all correlations are significantly different from zero at the .05 level or better. The fact that parents can agree on what their child is like to a better extent than either can agree with the teachers, indicates that the trouble is not just unreliability of parent ratings. Indeed, it is probably reasonable to assume that the interparent correlations would be as high as those for teachers if each parent had the same sample of child behavior to observe and rate. If one accepts this assumption, then it is necessary to also accept the conclusion that different information is contained in mother and father evaluations of the same factors on the same children, since the average parent correlations are significantly lower (at least the .01 level) than the average teacher correlations.

Additional Measures of Child Behavior

An additional four rating factors developed by Peterson as a part of this same project were available for relating to parent behavior. Peterson had parents and teachers rate the children on 20 Cattell-type bipolar scales (see Cattell, 1957, Appendix 1) which have fairly

² The parallel (uncorrected) reliabilities for the 59 nonparticipants were .70, .44, .70, .78, .83, and .83.

TABLE 8
INTERCORRELATIONS OF FACTOR SCORES

Teacher Factors							Mother Factors					Father Factors				
1	2	3	4	5	6		1	2	3	4	5	1	2	3	4	5
Teacher																
1	.80 ^a	-.31	-.47	.38	.52	-.66	.33	-.17	-.33	-.26	.24	.26	-.22	-.34	-.07	.40
2		.59	.51	.11	-.29	.11	-.20	.26	.19	.08	-.09	-.16	.23	.28	-.15	-.16
3			.73	.41	-.70	.28	-.04	.08	.30	.52	-.27	.04	.12	.24	.32	-.25
4				.70	-.20	-.49	.15	-.07	.04	.26	-.11	.04	.04	-.12	.25	.07
5					.90	-.67	.03	-.06	-.26	-.38	.40	.14	-.14	-.26	-.17	.39
6						.80	.00	.03	.12	.16	-.14	-.03	.09	.20	.04	-.26
Mother																
1							-.16	-.48	.21	.44	.59	-.28	-.34	.33	.33	
2								.44	.16	-.34	-.29	.34	.26	-.14	-.26	
3									.36	-.67	-.30	.30	.49	.15	-.39	
4										-.26	.18	.04	.13	.61	-.03	
5											.39	-.19	-.45	-.04	.56	
Father																
1												-.45	-.50	.51	.65	
2													.52	-.13	-.37	
3														-.14	-.67	
4															.17	

Note.—The factors named in terms of direction of high score are: 1. Hostile-Withdrawal, 2. Relaxed Disposition, 3. Lack of Aggression, 4. Submission, 5. Conduct Problem, and 6. Schoolroom Intelligence.
^a Diagonal values for teachers are the correlations between factor scores derived separately for Teacher 1 and Teacher 2 (i.e., before pooling).

precise behavioral descriptions defining end points. On the basis of a previous study, Peterson (unpublished) scored the 20 scales for the two largest factors, Adjustment and Extraversion. The Adjustment factor is briefly described by the following adjectives: patient, persevering, mannerly, good natured, calm, responsible, not jealous, cooperative, scrupulous, trusting. The Extraversion factor is described by: frank, happy-go-lucky, energetic, friendly, bold, cheerful, assertive, gregarious, composed, prefers companions of the opposite sex.

Peterson has also developed a 58-item problem checklist (as yet, unpublished) which both parents and teachers filled out. Two of the three factors from this checklist were included in the present analysis. They are Conduct Problem (disruptiveness, disobedience, attention seeking, fighting, irritability, destructiveness, etc.) and Personality Problem (inferiority feelings, anxiety, aloofness, reticence, depression, etc.). These four factors from Peterson's rating procedures were avail-

able for pooled parents and for pooled teachers.

As would be expected from the descriptions, there is considerable overlap between Peterson's child evaluations and those developed by the present writer. This overlap can be a virtue in that it allows one to check the degree to which a given finding is specific to one form of rating scale and therefore per-

TABLE 9
SUMMARY OF PARENT-TEACHER CORRELATIONS
OBTAINED BY PETERSON

Raters	Factors ^a				z to r Aver- age	Com- mon Vari- ance
	7	8	9	10		
Teacher 1-Teacher 2	.82	.68	.75	.64	.73	.53
Mother-Father	.48	.52	.62	.64	.57	.32
Parents-Teachers	.44	.24	.29	.26	.31	.10

^a Factor Definitions: 7. Conduct Problem, 8. Personality Problem, 9. Adjustment, and 10. Extraversion. The factors are numbered the same way as in Table 10 to facilitate cross-reference.

haps fortuitous. For example, the interparent, interteacher, and parent-teacher correlations (Table 9) on these four variables show practically an identical pattern to those just discussed from Table 8. The findings presented in Table 9 indicate that conclusion drawn from Table 8 is not restricted to the particular rating schedule used.

Factors among the Child Factor Scores

To simplify the interpretation of the dependencies among the child scores and to facilitate the later analysis of relationships to parent behavior, the 24 child factor scores were factored using the centroid method with varimax rotation. Table 10 presents the load-

ings of each child score on the four second-order factors. These four factors are readily interpreted as teachers' evaluations of *Personality Problem* behavior and *Conduct Problem* behavior, and parents' evaluations of the same two kinds of behavior. Factor scores for these second-order factors were computed by a simple addition of scores for the variables in Table 10 showing factor loadings which are not in parentheses.

THE ANALYSIS OF FACTORS IN PARENTAL RATINGS OF SELF AND EACH OTHER

One disappointment with our earlier work was with the Guilford inventories, particularly for fathers (Becker, Peterson, Hellmer,

TABLE 10
INTERDEPENDENCIES OF CHILD FACTOR SCORES

Variables	Varimax Factor Loadings			
	TPP ^a	TCP	PPP	PCP
Mother Ratings of Child				
1. Hostile-Withdrawal	(27)	-	(28)	(33)
2. Relaxed Disposition	-	-	-	-46
3. Lack of Aggression	-	-	-	-74
4. Submission	-	(-43)	50	-
5. Conduct Problem	-	-	-	64
Father Ratings of Child				
1. Hostile-Withdrawal	-	-	(30)	(39)
2. Relaxed Disposition	-	-	(-35)	-58
3. Lack of Aggression	-	-	-	-68
4. Submission	-	-	60	-
5. Conduct Problem	-	-	-	65
Pooled Parent Ratings				
7. Peterson Conduct Problem	-	(39)	-	68
8. Peterson Personality Problem	-	-	64	-
9. Peterson Adjustment	-	-	-	-72
10. Peterson Extraversion	-	(32)	-55	-
Pooled Teacher Ratings				
1. Hostile-Withdrawal	74	(42)	-	-
2. Relaxed Disposition	-	-52	(-32)	-
3. Lack of Aggression	-	-89	-	-
4. Submission	79	(-42)	-	-
5. Conduct Problem	-	80	-	-
6. Schoolroom Intelligence	-71	(-36)	-	-
7. Peterson Conduct Problem	-	86	-	-
8. Peterson Personality Problem	80	-	-	-
9. Peterson Adjustment	-	-89	-	-
10. Peterson Extraversion	-90	-	-	-

^a The abbreviations stand for Teacher Personality Problem (TPP), Teacher Conduct Problem (TCP), Parent Personality Problem (PPP), and Parent Conduct Problem (PCP).

Shoemaker, & Quay, 1959). When the two main child problem factors (Personality Problem and Conduct Problem) were examined for relationships to mother's and father's behavior, only three Guilford variables were found to load .30 or higher (.30, .31, and .32). Examination of the correlations between the 13 Guilford factors and 6 child variables indicated that only 2.6% of the father correlations were significant at the .05 level, while 18% of the mother correlations were significant. The mother correlations mainly related mother anxiety and hostility to child conduct problems. Considering the fact that groups were extremes on adjustment variables (clinic versus well adjusted families), the magnitude of the correlations (under .40) was not impressive. These findings, coupled with a paucity of associations between the Guilford scales and factors in the Fels Parent Behavior Scales, led to the belief that the generalized self-appraisal factors measured by the Guilford inventories have little to do with what parents do (or say they do) as parents.

To further explore the value of self-ratings, the hypothesis was generated that self-rating factors more directly associated with the *parental role* would show more relationships to child behavior. A rating schedule was devised to test this hypothesis (informally) and at the same time to systematically examine interparental perceptions³ in relationship to child behavior.

The Parent Rating Schedule

The rating schedule consisted of 73 bipolar seven-point rating scales with antonym pairs of adjectives defining the extremes. Scales were selected to sample (a) the factors represented in the Fels Parent Behavior Rating Scales (Roff, 1949), (b) Cattell's personality

factors (Cattell, 1957), and (c) the factors found by Osgood and Becker (unpublished) using the semantic differential to assess the self-concept. The more unusual adjective pairs, such as *real-unreal*, are from this latter source. For each of the 73 variables, parents were asked to rate the following concepts (father variations are in parentheses):

1. My-self-in-relationship-to-my-child
2. My-self-in-relationship-to-my-husband (wife)
3. My-husband (wife)-in-relationship-to-my-child
4. My-husband (wife)-in relationship-to-me
5. The-person-I-would-like-to-be-in-relationship-to-my-child
6. The-person-I-would-like-to-be-in-relationship-to-my-husband (wife)

The instructions, besides explaining the general nature of the scales, encouraged the parents to think carefully about the *particular* relationship before making the ratings. Parents were also advised to beware of halo-type biases (not stated in these terms, of course) and that the results would be confidential. Finally, they were instructed not to go back once a rating had been made.

The average time for the ratings was 40–60 minutes. Typically, fathers made their ratings at home after a 3-hour interview and testing session which the complete project involved. Mothers made their ratings during a second visit to the clinic. Parents were asked not to discuss ratings with one another until both had finished. There is no reason to doubt that this request was not followed.

Factor Analysis of the Rating Schedule

In a previous pilot study, Silverstein (1958) had demonstrated the feasibility of using a single factor-scoring system to score each relationship concept for each parent. A single scoring system is necessary if cross-concept comparisons are to be made. With the aid of an electronic computer, correlations were computed by summing over Concepts 1–4 (listed above) and over persons (both mothers and fathers). Each correlation entering the factor analysis was based on ratings from 62 mothers and 60 fathers over four concepts or 488 elements in all. The ideal relationship concepts were left out because it was not believed they would have enough variance to make computation of factor scores worth-

³ Usage of the terms perception, attitude, and behavior needs clarification. Ratings can reflect concepts or attitudes of the rater as well as the behavior of the person being rated. With only one rater, it is not always possible to tell which effect one is dealing with. We have tended to call our variables *ratings* or *perceptions* if no other information is available, *self-attitudes* if the ratings are indiscriminately related to other variables, and *behaviors* if there is consensual agreement with another rater. On occasion our choice of terms has had to be arbitrary so as not to encumber communication.

TABLE 11

DESCRIPTION OF THE VARIMAX FACTORS IN PARENTAL RATINGS OF SELF AND EACH OTHER

Loading	Description	Loading	Description
Factor 1—Hostile-Withdrawal			
77	Loving.....unloving	45	Not-prone-to-anger.....prone to anger
68	Responsive.....aloof	41	Rational.....arbitrary
68	Warm.....cold	38	Conscientious.....conscienceless
67	Emotionally close.....detached	Factor 5—Commonness (low IQ?)	
63	Devoted.....rejecting	66	Formed.....formless
63	Forgiving.....begrudging	62	Individualistic.....ordinary
61	Sociable.....unsociable	54	Curious.....uninquiring
53	Approving.....disapproving	51	Deep.....shallow
51	Kind.....cruel	48	Flexible.....inflexible
50	Cooperative.....hostile	45	Adventurous.....timid
49	Meaningful.....meaningless	45	Quick.....slow
48	Vigorous.....inert	39	Real.....unreal
46	Outgoing.....self-centered	34	Effective.....ineffective
44	Soft hearted.....hard hearted	Factor 6—Solicitousness	
Factor 2—Dominance-Strictness		59	Nonchalant.....anxious
71	Permissive.....strict	43	Not jealous.....jealous
64	Mild.....severe	40	Underhelping.....overhelping
63	Submissive.....dominant	40	Inactive.....active
63	Democratic.....authoritarian	39	Lax.....vigilant
60	Nondemanding.....demanding	37	Objective.....emotional
57	Pliable.....rigid	37	Nonsuggesting.....suggesting
56	Nonrestrictive.....restrictive	32	Trusting.....suspicious
53	Uncritical.....critical	Factor 7—Nonprotectiveness	
50	Sensitive.....tough	63	Sheltering.....exposing
50	Easy going.....irritable	53	Protective.....nonprotective
49	Soft.....hard	Factor 8—Harmony	
46	Weak willed.....strong willed	46	Contentious.....concordant
Factor 3—Nervousness		42	Punitive.....nonpunitive
71	Confident.....unsure	39	Conflicted.....nonconflicted
67	Clear.....confused	38	Taking.....giving
64	Relaxed.....tense	36	Threatening.....reassuring
64	Happy.....depressed	Factor 9—Social Effectiveness	
60	Fearless.....fearful	48	Disorganized.....organized
53	Energetic.....fatigued	48	Colorless.....colorful
48	Free from guilt.....guilt ridden	47	Unsuccessful.....successful
43	Calm.....excitable	37	Boring.....interesting
38	Optimistic.....pessimistic	Factor 10—Playfulness	
Factor 4—Immature, Aggressive Emotionality		55	Serious.....playful
54	Level.....fluctuating	48	Humorless.....humorous
54	Patient.....impatient	38	Thwarts curiosity.....satisfies curiosity
51	Consistent.....inconsistent		
46	Understanding.....not understanding		

while. This belief later proved to be wrong. After computing correlations, 15 centroid factors were extracted using fixed-unit communalities. Ten factors, accounting for 50% of the total variance,⁴ were then rotated to or-

⁴ When fixed-unit communalities are used, the total common variance is markedly overestimated and the percentage of common variance is underestimated.

thogonal simple structure using the varimax criterion (Kaiser, 1958).

The varimax rotated factors have been deposited with ADI.⁵ Because each variable

⁵ The rotated varimax factors have been deposited with the American Documentation Institute. Order Document No. 6467 from ADI Auxiliary Publications Project, Photoduplication Service, Library of

TABLE 12
SPLIT-HALF FACTOR SCORE RELIABILITIES

Concept	Factors ^a									
	1	2	3	4	5	6	7	8	9	10
Mother										
1. Self-child	82	82	49	54	78	69	66	26 ^b	72	45
2. Self-spouse	94	79	91	66	74	71	54	17	76	69
3. Spouse-child	92	88	88	76	83	81	79	48	70	66
4. Spouse-self	93	80	82	80	81	63	68	13	72	74
5. Ideal-child	72	64	73	62	70	29	74	28	68	63
6. Ideal-spouse	83	72	81	75	77	25	77	43	67	58
Father										
1. Self-child	87	88	85	82	80	52	66	52	59	69
2. Self-spouse	92	84	74	78	84	69	65	47	70	75
3. Spouse-child	93	82	86	84	84	56	78	25	52	83
4. Spouse-self	94	87	81	85	87	62	80	29	83	64
5. Ideal-child	88	76	61	76	87	40	62	36	76	70
6. Ideal-spouse	94	86	73	82	90	15	73	15	62	74

^a Factor titles are: 1. Hostile-Withdrawal; 2. Dominance-Strictness; 3. Nervousness; 4. Immature, Aggressive Emotionality; 5. Commonness (Low IQ?); 6. Solicitousness; 7. Nonprotectiveness; 8. Harmony; 9. Social Effectiveness; 10. Playfulness.

^b Italicized variables were eliminated from further analyses.

loaded at least one factor higher than .29, a complete listing of the variables is represented in the description of factors given in Table 11. In preparing Table 11, variable descriptions have been reversed where necessary to make all loadings positive. The original ordering was random with regard to direction. It should be noted that Table 11 contains only the variables used in *scoring* each factor. The label given a factor indicates the direction of a high factor score. Factor titles have been given to facilitate communication. The reader familiar with the literature will see many similarities between the present factors and those reported previously by Cattell (1957) and Roff (1949) in behavior ratings.

Computation of Factor Scores

Factor scores were computed by the method of standard score addition of salient variables, weighted according to factor loading. In computing factor scores no variable was scored for more than one factor. This procedure avoided the building in of spurious correlations among the factors due to corre-

lated errors. As noted earlier, even though an orthogonal solution was used and spurious correlations were avoided, it is to be expected that the factor scores *derived by the method outlined above* will tend to be correlated to the extent that the factors actually required an oblique solution.

Factor Score Reliability

To find out whether the factors derived by summing over parents and over concepts are applicable for both parents and for the various concepts, two split-half estimates were made of each factor score, and factor consistency coefficients were computed. Table 12 presents a summary of these correlations as corrected by the Spearman-Brown prophecy formula to give estimates of the reliability of the full factor scores. As in the pilot study by Silverstein (1958), the reliabilities were generally found to be quite high. On the basis of the results given in Table 12, Factor 8 (Harmony) was eliminated completely, and Factor 6 (Solicitousness) was eliminated for the ideal-relationship concepts. Factors 3 (Nervousness) and 10 (Playfulness) had reliabilities below .50 for the self-in-relationship-to-child ratings by mothers. To maintain symmetry between the mother and father matrices

(and thus facilitate computations), these latter two variables were not eliminated.

Relationships among the Parent Factor Scores

The correlations among the parent scores were computed in two 52 variable matrices covering factors for the following relationship concepts.

MATRIX A		MATRIX B	
Father 1. Self-child		Mother 1. Self-child	
Mother 3. Husband-child		Father 3. Wife-child	
Father 2. Self-wife		Mother 2. Self-husband	
Mother 4. Husband-me		Father 4. Wife-me	
Father 5. Ideal-child		Mother 5. Ideal-child	
Father 6. Ideal-wife		Mother 6. Ideal-husband	

Since the patterns of interdependencies tended to repeat, the structure of the 2,652 correlations in Matrices A and B can be summarized by two small tables.

Table 13 presents the median correlations between relationship concepts taken over the eight or nine factor scores for each concept. Two sets of comparisons are of particular interest in Table 13. First, the comparison between self-ratings and spouse-ratings in relationship-to-child or in relationship-to-spouse indicated fair agreement for some factors. The median correlations ranged from .35 to .44. Secondly, the comparisons of parental ratings of self or of each other in the two role situations (with child or with spouse) showed median correlations which ranged from .54 to .70. Since the median reliabilities are mostly in the low .80s, considerable room remains for differential correlations of child versus parent-in-relationship-to-spouse variables.

TABLE 13

MEDIAN CORRELATIONS BETWEEN RELATIONSHIP CONCEPTS OVER FACTORS

Relationship Concepts	Matrix A				
	M3	F2	M4	F5	F6
Father 1. Self-child	35	60	25	21	21
Mother 3. Husband-child		26	58	05	06
Father 2. Self-wife			42	10	18
Mother 4. Husband-me				-09	04
Father 5. Ideal-child					40
Father 6. Ideal-wife					

Relationship Concepts	Matrix B				
	F3	M2	F4	M5	M6
Mother 1. Self-child	36	54	33	38	06
Father 3. Wife-child		33	70	00	08
Mother 2. Self-husband			44	27	12
Father 4. Wife-me				00	03
Mother 5. Ideal-child					27
Mother 6. Ideal-husband					

Table 14 presents the median correlations among the parent factor scores across all 12 relationship concepts (i.e., both Matrix A and B). As was apparent in the original factor analysis, there is a second-order factor among the parent factors defined best by the correlations between the factors labeled Hostile-Withdrawal; Nervousness; Immature, Aggressive Emotionality; Commonness; and Social Ineffectiveness. The second-order factor might be interpreted as a halo effect, an evaluative factor, or a social desirability effect. Whether the factor is looked on as an undesirable bias or an important aspect of personality depends on one's purposes. From

TABLE 14
MEDIAN CORRELATIONS BETWEEN FACTORS OVER RELATIONSHIP CONCEPTS

Factor Title	2	3	4	5	6	7	9	10
1. Hostile-Withdrawal								
2. Dominance-Strictness	44	61	68	62	-36	29	-71	-59
3. Nervousness		32	43	16	-04	19	-27	-32
4. Immature Emotionality			64	56	-10	15	-67	-40
5. Commonness				48	-11	18	-60	-30
6. Solicitousness					-32	16	-65	-52
7. Nonprotectiveness						-34	25	26
9. Social Effectiveness							-24	-26
10. Playfulness								52

the point of view of the writer, the factor will be considered desirable if it correlates with *independent* evaluations of child behavior and a nuisance if it does not.

Additional Derived Scores

Two additional kinds of scores were derived from the parent factor scores. First D^2 scores were computed between parallel sets of parental rating factors: e.g., the difference (squared) between father's ratings of self-in-representation-to-child and mother's ratings of husband-in-representation-to-child. Interpersonal theory would lead to the prediction that the greater the discrepancies in interparental perceptions, the more likely the presence of problem behavior in children.

Scores were also obtained for the second-order evaluative factors among the parental ratings. Evaluative factors were computed for the parent-in-representation-to-child concepts and for the ideal-representation concepts on the basis of a factor analysis of the parent factor scores. To the extent these second-order factors represent more than a specific rating bias (and there is evidence to this effect), they will permit an examination of the relationship of generalized positive or negative parental attitudes to child behavior.

THE RELATIONSHIP BETWEEN THE PARENT AND CHILD FACTORS

As a guide in examining the results, several hypotheses were formulated from our previous research (Becker, et al., 1959; Peterson, Becker, Hellmer, Shoemaker, & Quay, 1959) in addition to those which have been discussed in previous sections of this report. Hypotheses will be presented as data relevant to their tenability is examined.

The data to be evaluated consist of 2,496 correlations between the 24 child measures and 104 parent measures, the correlations of the 4 second-order child measures with the parent measures, and the correlations of the child measures with the D^2 scores and the second-order evaluative factors. Fortunately there are several ways of dealing with the quantity of data which will minimize capitalization on chance relationships. The correlations between the 24 child measures and the 104 parent measures will be discussed first.

Significance of Correlations by Blocks of Variables

In Table 15, the percentage of correlations significant at the .05 level has been compiled for the various submatrices interrelating the 104 parent and 24 child variables. Table 15 permits one to appraise the probable significance of relationships between parent and child variables. However, interdependencies among the variables and correlated sampling errors prohibit precise probability statements. Table 15 serves as a guide for further elimination of blocks of variables, permits a test of some hypotheses, and reveals other facts of interest.

A large and consistent rating bias is apparent in the father correlations such that if he rates himself favorably, he also rates his wife and child favorably on most factors. This leads to a high percentage of significant correlations among all father ratings. The percentages of significant correlations between father ratings of child and father ratings of self and spouse are 84%, 47%, 73%, and 56%. When *mother* ratings of child behavior are compared with the same father ratings of self and spouse, the percentages are 53%, 16%, 27%, and 33%. Thus, there is an average drop of 33% in significant correlations when rater contamination is avoided. A similar, but less marked, effect can also be demonstrated in the mother ratings. Obviously, it is necessary to control for this effect in analyzing the relationships between parent and child behavior.

Support for two of the hypotheses guiding this research is gained from Table 15. First it was hypothesized that self-ratings more directly associated with the parental role would show more correlations with child behavior. While a direct comparison with the previous findings using the Guilford inventories is not possible, there are some comparisons *within* the present study which support the hypothesis. Using only correlations uncontaminated by rater bias, the percentage of significant correlations between child behavior and parent ratings of self-in-representation-to-child (F1, M1) was found to be 24% for fathers and 9% for mothers. The corresponding percentages for self-in-representation-to-spouse were

TABLE 15

THE PERCENTAGE OF SIGNIFICANT CORRELATIONS (.05 level) BETWEEN PARENT AND CHILD VARIABLES ANALYZED BY CONCEPT AND BY CHILD RATER

Parent Relationship Concept	Child Variables by Rater					Row Means ^b	
	Teacher B(6) ^a	Teacher P(4)	Mother B(5)	Father B(5)	Parents P(4)		
Father 1. Self-child (9)	07	17	53	84	53	42	35
Mother 3. Husband-child (9)	15	14	51	31	30	28	
Father 2. Self-wife (9)	02	00	16	47	19	17	14
Mother 4. Husband-me (9)	04	06	33	11	08	12	
Mother 1. Self-child (9)	06	06	42	16	28	19	24
Father 3. Wife-child (9)	00	00	27	73	55	30	
Mother 2. Self-husband (9)	06	00	36	00	14	11	18
Father 4. Wife-me (9)	00	03	33	56	30	24	
Father 5. Ideal-child (8)	06	06	20	13	25	14	11
Mother 5. Ideal-child (8)	04	03	25	08	06	09	
Father 6. Ideal-wife (8)	00	00	10	10	00	04	6
Mother 6. Ideal-husband (8)	00	00	25	03	13	08	
Column Means	04	05	31	30	24	18.6	

^a B stands for Becker's scores and P for Peterson's scores. By multiplying the numbers in parentheses after each title, one can determine the number of correlations in each submatrix upon which the percentage is based.

^b The average of the two adjacent row means is given in the right hand column.

6% and 2%. While these figures are not representative of the general level of significance of the correlations, because the relatively low teacher correlations are weighted double, they clearly indicate that more significant correlations are found when parents rate their behavior specifically in relation to the child under study. These same effects also hold for ratings by spouse, i.e., mother's ratings of father and father's ratings of mother-in-relation-ship-to-child and for the ideal relationships.

The second hypothesis to gain support from Table 15 concerns the relative significance of father's attitudes and behavior in relationship to mother's. On the basis of previous work, it was expected that the number of associations between father and child behavior would equal or exceed that for mothers. Table 15 reveals that 35% of the correlations between father-in-relation-ship-to-child variables and child behavior were significant, while only 24% of the corresponding mother correlations were significant. When only the uncontaminated correlations are considered, the figures

are 22% (60/270) for fathers and 9% (24/270) for mothers. Again, these latter figures overweigh the low teacher correlations.

A further reduction in the number of correlations to be examined in detail can be accomplished by a few summary statements. The reader will find the structural layout of Table 15 helpful in understanding the following points:

1. When the F2, F4, F6 and the M2, M4, M6 variables did correlate significantly with child behavior, typically the same variables were involved as were significant for the F1, F3, F5 and the M1, M3, and M5 variables. Since the former correlations were generally fewer and lower, and since they added no new information to the prediction of child behavior, they will not be discussed further.

2. By eliminating from the remaining submatrices, those blocks of variables where the percentage of significant correlations is less than 10%, as well as those subject to rater bias contamination, the number of correlations to be examined is reduced to 346.

3. If the correlations of the four second-

order child factors are included for those parent variables surviving the above eliminations, the total number of correlations to be given close attention is 459. Of these 459 correlations, 122 or 26.6% reached the .05 level of significance. These correlations are presented in Table 16.

Evaluation of Relationships between Parent and Child Factors

Space does not permit a complete descriptive elaboration of the implications of the correlations in Table 16. As noted in Table 10, Child Factors 2, 3, and 5 load primarily on a second-order Conduct Problem factor and

TABLE 16
SIGNIFICANT CORRELATIONS BETWEEN PARENT AND CHILD FACTOR SCORES

		Child Scores ^a									
Rater		Mother					Parents		Teachers		
Father 1. Self-Child		1	2	3	4	5	7	8	7	8	
1. Hostile-Withdrawal		.46	-.31	-.34	-	.47	.53	.28	-	-	
2. Dominance-Strictness		.29	-.30	-.26	-	.24	.34	-	-	-	
3. Nervousness		.36	-	-.30	-	.31	.44	.24	.30	-	
4. Aggressive Emotionality		.23	-	-.35	-	.36	.43	-	-	-	
5. Commonness (low IQ?)		.29	-	-	-	-	.30	.31	-	-	
6. Solicitousness		-.30	-	-	-	-.27	-.26	-.23	-	-	
7. Nonprotectiveness		-	-	-	-	-	-	-	-	-	
9. Social Effectiveness		-.38	.29	.26	-	-.35	-.43	-.31	-	-	
10. Playfulness		-.48	.23	-	-	-.31	-.42	-.37	-	-	
Rater		Father					Parents		Teachers ^b		
Mother 3. Husband-Child		1	2	3	4	5	7	8	7	8	
1. Hostile-Withdrawal		.45	-	-	.27	.35	.39	.23	-	-	
2. Dominance-Strictness		.30	-	-	-	.33	.41	-	.32	.25	
3. Nervousness		.36	-	-.39	-	.40	.44	-	.38	-	
4. Aggressive Emotionality		-	-	-	-	.26	.40	-	-	-	
5. Commonness (low IQ?)		.29	-	-	.35	-	.24	.25	-	-	
6. Solicitousness		-	-	-	-	-	-	-	-	-	
7. Nonprotectiveness		-	-	-	-	-	-	-	-	-	
9. Social Effectiveness		-.36	-	.26	-.33	-.27	-.38	-.27	-	-	
10. Playfulness		-	-	-	-	-	-	-	-	-	
Rater		Father					Parents		Teachers		
Mother 1. Self-Child		1	2	3	4	5	7	8	7	8	
1. Hostile-Withdrawal		.28	-	-.24	-	.25	.42	-	-	-	
2. Dominance-Strictness		-	-	-	-	-	.24	-	.24	-	
3. Nervousness		-	-	-.39	-	.23	.54	-	-	-	
4. Aggressive Emotionality		-	-	-	-	-	.44	-	-	-	
5. Commonness (low IQ?)		-	-	-	-	-	.28	-	-	-	
6. Solicitousness		-	-	-	-	-	-	-	-	-	
7. Nonprotectiveness		-	-	-	-	-	-	-	-	-	
9. Social Effectiveness		-	-	.30	-	-.28	-.53	-	-	-	
10. Playfulness		-	-	-	-	-	-	-	-	-	

Note.—Correlations of .23 are significant at the .05 level. Correlations of .31 are significant at the .01 level.

^a Child factors are identified as follows: 1. Hostile-Withdrawal; 2. Relaxed Disposition; 3. Lack of Aggression; 4. Submission; 5. Conduct Problem; 6. Schoolroom Intelligence; 7. Conduct Problem, second-order factor; 8. Personality Problem, second-order factor.

^b The 54 correlations between teachers' evaluations of the children and mothers' evaluation of father have been omitted to facilitate tabular presentation. The eight significant correlations in this submatrix were: (a) father Hostile-Withdrawal correlated .23 with Child Factor 1; (b) father Dominance-Strictness correlated .31 with Child Factor 1, -.27 with Child Factor 2, -.26 with Child Factor 3, and .23 with Child Factor 5; (c) father Nervousness correlated -.29 with Child Factor 2, -.34 with Child Factor 3, and .34 with Child Factor 5.

TABLE 16—Continued

Rater	Mother					Parents		Teachers	
Father 3. Wife-Child	1	2	3	4	5	7	8	7	8
1. Hostile-Withdrawal	50	-	-23	-	25	46	36	-	-
2. Dominance-Strictness	-	-	-	-	-	-	-	-	-
3. Nervousness	34	-	-	-	-	37	-	-	-
4. Aggressive Emotionality	37	-	-	-	-	39	25	-	-
5. Commonness (low IQ?)	36	-	-	-	-	32	34	-	-
6. Solicitousness	-35	-	-	-	-	-	-31	-	-
7. Nonprotectiveness	-	-	-	-	23	29	-	-	-
9. Social Effectiveness	-48	-	-	-	-32	-44	-	-	-
10. Playfulness	-48	-	-	-25	-	-31	-43	-	-

Rater	Mother					Parents		Teachers	
Father 5. Ideal-Child	1	2	3	4	5	7	8	7	8
1. Hostile-Withdrawal	-	-	-	-	-	-	-	-	-
2. Dominance-Strictness	-	-23	-	-31	-	-	-	-	-
3. Nervousness	29	-	-28	-	-	-	-26	30	-
4. Aggressive Emotionality	-	-	-28	-	-	27	-	-	-
5. Commonness (low IQ?)	-	-	-24	-	-	-	-	-	-
7. Nonprotectiveness	-	-	-	-	-	-	-	-	-
9. Social Effectiveness	-28	-	-	-	-	-	-	-	-
10. Playfulness	-	-	25	-	-	-	-	-	-

Child Factor 4 loads primarily on a Personality Problem factor, while Child Factor 1 loads on both. Because our hypotheses were primarily focused on these second-order factors, the discussion of Table 16 will be primarily restricted to the second-order child factors.⁶ The interested reader will find it instructive to examine more closely the relationships between the parent factors and first-order child factors.

Child Conduct Problem. On the basis of previous work (Becker, et al., 1959; Peterson, et al., 1959), it was hypothesized that child conduct problems are related to general maladjustment, open venting of negative emotions, and arbitrary and inconsistent discipline by both parents. The father-child correlations in Table 16 indicate that the aggressive, uncontrolled, hard to discipline child (as seen by parents) has a father who shows more Hostile-Withdrawal (cold, aloof, re-

jecting); Dominance-Strictness (severe, authoritarian); Nervousness (unsure, confused, tense); Immature, Aggressive Emotionality (fluctuating, impatient, inconsistent, prone to anger); Social Ineffectiveness (disorganized, unsuccessful); and Seriousness. The picture is much the same whether mother evaluates father or father evaluates himself. Only the Nervousness and Dominance-Strictness factors correlate significantly with teachers' evaluations of Conduct Problem. The mother correlations with child Conduct Problem are very similar to those for father, suggesting a picture of rejection, withdrawal, nervousness, inconsistency, and irritable aggressiveness. The primary difference between the mother and father correlations with child Conduct Problem is that the Dominance-Strictness factor appears to play less of a role for mothers. In all the pattern of correlations strongly supports the hypothesis.

Following the general behavior theory developed by the Iowa-Yale group (Dollard & Miller, 1950; Sears, et al., 1957; Whiting & Child, 1953), Bandura and Walters (1959) have hypothesized that frustration arising from a lack of affectional nurturance and a

⁶ Although these second-order factors lead to some confounding of rater bias effects when correlated with parent behavior factors, it is possible to estimate the degree of such influence by examining the uncontaminated correlations of the first-order child factors with the parent behavior factors.

punitive attitude on the part of at least one parent are essential conditions for the development of aggressive behavior. In addition, they consider inconsistency in disciplinary practices and the kind of model the parents present as critical factors. If the father and mother markedly differ in demands and expectations, or if they are inconsistent in these areas, the child is left without clearly defined standards of behavior. Bandura and Walters (1959) find strong support for their position when extreme groups of aggressive and non-aggressive adolescents are studied using primarily the structured interview methods developed by Sears et al. (1957). It is apparent that the present findings based on self-ratings and cross-parental ratings lead to a similar conclusion. Probably most impressive is the fact that the behavior theory hypotheses appear to stand up even when the behavior of most of the children falls within normal limits.

Child Personality Problem. Our empirical hypothesis concerning child Personality Problem predicted associations to measures of paternal maladjustment and autocracy and a relative independence of maternal behavior. The results for fathers tend to support only the maladjustment component of the hypothesis. Both mother and father agree that fathers of Personality Problem children show more Hostile-Withdrawal (aloof, cold, rejecting), more Commonness (formless, uninquiring, inflexible, timid), and more Social Ineffectiveness (disorganized, colorless, unsuccessful). Fathers add to picture associations with Nervousness, Nonsolicitousness, and Seriousness. Actually the father picture is one of *general social withdrawal* not unlike the Personality Problem pattern found in the child. No associations are found between mothers' ratings of self and child Personality Problem. However, when father rates mother, several significant relationships to child Personality Problem are found. In fact, when father considers mother most things *bad* the child is rated more in the Personality Problem direction.

There are at least two possible explanations of the differences between mother and father ratings of mother in relationship to child Personality Problem. First, the results might reflect *father attitude* rather than mother be-

havior. However, correlations between the second-order evaluative factors (discussed later) indicate that mother tends to agree ($r = .46$) with father's evaluation of her. The second possibility is that a covertly rejecting mother defends from awareness her negative behavior which the father clearly sees. This interpretation fits with findings by Hewitt and Jenkins (1946) that overinhibited children came from homes where father's discipline was inconsistent; where father was hypercritical and unsociable; and where mother was unsociable, dominating, and *covertly rejecting*.

From animal studies of conflict and punishment (see Diamond, 1957) and from the work of the Iowa-Yale group, personality problems (dependence, anxiety, and withdrawal) are to be expected theoretically where sufficient warmth and acceptance is provided to promote a strong dependency need, and where punishment for misdeed is strong and inconsistent, and where withholding of love is used as a means of discipline. Unfortunately, most of this theory has not been developed in a context of research where *both* parents were considered, so the differential relevance of mother and father is unclear from theory. The work of Hewitt and Jenkins (1946), the previous study by Becker et al. (1959), and the present results all suggest that maladjustment of the father as manifested by hostility and social withdrawal is important in the development of shy and anxious children.⁷

Father Ideal Relationship to Child. The ideal correlations in Table 16 were rather surprising, since one would expect ratings of an ideal concept to cluster around a social norm with little variance. That the ideal concept factors did have appreciable variance was forewarned by the reliability findings. The correlations given in Table 16 indicate that the variance in ideal ratings is related to child behavior in important ways. There are two interesting things to note about these cor-

⁷ Initial results from our interview rating data suggest further confirmation of this. Ratings of a hostile relationship of father to child correlate significantly with both Conduct Problems (.36) and Personality Problems (.31) when Peterson's problem checklist is used to obtain the latter measures. The corresponding correlations with mother's hostility were .37 and .04.

relations. First, because the F5 factors correlate very lowly with the F1 factors (median .21) and the M3 factors (median .05), new information relating to the prediction of child behavior is contained in these relationships. For example, with a simple equal-weight addition of F1 and F5 scores, the correlations with child behavior are increased for 10 of the 24 significant variables in the submatrix based on father's ratings of self (F1) and mother's ratings of child. Secondly, since the direction of the correlations is the same as that found for self-in-relationship-to-child correlations, these relationships would not have been found if the ideal ratings had been used in the typical way to compute discrepancy scores.

The nature of the correlations between father's ideal-in-relationship-to-child and child behavior are all in a direction to indicate that if father's conception of his ideal relationship is more loving, democratic, and emotionally mature, the child is rated as being better adjusted, outgoing, less demanding, etc.

Relationship of Interparental D² Scores to Child Second-Order Factors

The factor score discrepancies between mother's and father's perception of father-in-relationship-to-child showed significant rela-

tionships to the second-order child Conduct Problem factors as evaluated by parents ($r = .31$) and by teachers ($r = .27$). The similar correlations for father-in-relationship-to-wife were .31 and .30. The relationships to Personality Problems for the above discrepancies were insignificant. When the discrepancies concerned the perception of mother's relationship to child or to husband, the only significant correlation was between the mother-in-relationship-to-child discrepancy and parental ratings of Personality Problem ($r = .23$). This latter correlation should not be taken too seriously without cross-validation since only one of the four possible relationships to child Personality Problem was significant. On the other hand the relationships to child Conduct Problem were significant for all four comparisons made. If interparental consistency in discipline and lack of conflict can be assumed to be related to small discrepancies in perception, then these findings fit well with the behavior theory account of the etiology of aggressive behavior.

Relationship of Parent Evaluative Factors to Child Scores

The correlations between the parental second-order evaluative factors and some of the

TABLE 17
CORRELATIONS BETWEEN PARENTAL SECOND-ORDER EVALUATIVE FACTORS AND CHILD SCORES

ORDER EVALUATIVE FACTORS AND CHILD SCORES								
Rater		Child Scores ^a						
Father Evaluative Factors ^b		Mother					Parents	
		1	2	3	4	5	7	8
F1. Self-Child		-48	29	36	-09	-44	-55	-33
F3. Wife-Child		-48	16	22	-13	-40	-46	-34
F5. Ideal-Child		-21	25	30	18	-08	-24	02
F6. Ideal-Wife		-19	-04	16	16	-06	-11	03
Rater		Father					Parents	
Mother Evaluative Factors		1	2	3	4	5	7	8
M1. Self-Child		-26	08	32	-12	-26	-53	-13
M3. Husband-Child		-38	-02	28	-26	-34	-42	-18
M5. Ideal-Child		-08	-13	-05	-10	03	02	-12
M6. Ideal-Husband		-21	10	14	-19	-09	-08	-27

^a See Table 16 for description of child scores.

^a See Table 16 for description of child variables.
^b High score indicates a positive evaluation.

child scores are presented in Table 17. Because rater bias contamination slightly inflates the correlations between the evaluative factors and the child second-order factors, correlations with uncontaminated first-order child factors have also been included to permit the reader to judge the effect of this inflation. It is apparent that the general goodness of father evaluation of self- or wife-in-representation-to-child is related to a lack of both Personality and Conduct Problems in the child. Similar evaluations by mother are related only to a lack of Conduct Problems. The relationships of the evaluative factors in the ideal-relationship ratings to child behavior confirm findings discussed earlier, except that child Personality Problem is now found to be significantly associated with mother having a more negative ideal of what she would like to be in relationship to her husband. In view of the lack of significant correlations of this mother-ideal factor with uncontaminated child scores, this finding should be viewed with caution.

The reader is left to his own devices to pull out some of the fuller implications of the correlations in Table 17. One implication, apparent when the appropriate evaluative factor is partialled out of the correlations in Table 16, is that the second-order factors account for a large part of the covariance with child behavior. This finding tends to emphasize again the importance of the general parental attitude of warmth or acceptance. One final set of correlations should be of interest. The correlation between mother's evaluation of father-in-representation-to-child and father's evaluation of himself was .40. The parallel correlation for mother-in-representation-to-child was .46. These correlations indicate that the general evaluative factors are more than *just* attitudes, but also reflect behavior observable by the other parent.

Methodological Problems

The present findings give no information about causality. It is just as possible that child behavior is generating parental reactions as the contrary, or it is also possible that similar genotypes are the basis for some of the found covariations. The present methodology is defensible only as a necessary first

step toward finding truly causal relationships.

The rather large *difference* in number of significant associations to mother and father behavior was somewhat surprising. Such a striking difference does not seem reasonable and the question of artifact must be considered. If one assumes that father's ratings of mother and child are less valid than father's ratings of self or mother's ratings of self, father, and child, then the obtained results are explicable. This hypothesis is plausible since most fathers do not spend much time at home when mother and child are interacting. Further support for this hypothesis is found in the fact that father's ratings seemed to suffer more from a general bias effect than did mother's ratings. Such an effect is most expected when ratings are made with insufficient information available. Perhaps other data now under analysis will shed light on this problem.

SUMMARY

Three problems were investigated: (a) the interrelationships of factors in mothers', fathers', and teachers' ratings of the same children; (b) the nature of and interrelationships of factors in parental ratings of self and each other; and (c) the relationship between the factors obtained in *a* and *b* above. The primary sample consisted of 71 kindergarten children and their parents.

Problem *a* was investigated by having mothers, fathers, and two teachers rate the children on a 72-variable bipolar adjective rating schedule. Two other kinds of rating instruments were also used in evaluating child behavior. Problem *b* was investigated by having parents rate the following concepts on a 73-variable bipolar rating schedule: myself-in-representation-to-my-child, myself-in-representation-to-my-spouse, my-spouse-in-representation-to-my-child, my-spouse-in-representation-to-me, the person-I-would-like-to-be-in-representation-to-my-child, and the person-I-would-like-to-be-in-representation-to-my-spouse.

Highlights of the main findings are as follows:

1. While it was possible to find reasonable matchings between mothers', fathers', and teachers' ratings of the children in terms of factor loading patterns, the parent and

teacher scores for children on the same factors showed little correlation. For five factors, the average correlation was .76 between teachers, .52 between parents, and .34 between parents and teachers. These findings indicated a need to keep parent and teacher rating measures separated in developing descriptions or testing explanations of child behavior.

2. When parents were asked to rate their behavior *with the child under study*, rather than with spouse, many more significant associations with child behavior were found. For example, using only correlations uncontaminated by rater bias, 24% of the correlations between ratings of father-in-relationship-to-child and child measures were significant, while the similar figure for ratings of father-in-relationship-to-wife was 6% significant correlations. These results emphasized the importance of focusing directly on parental-role behavior rather than general personality in seeking relationships between parent and child behavior.

3. The relationships of parent rating factors to child behavior were significant, by and large, only when the parents also rated the children. Associations to teachers' ratings were few, being mainly with teachers' evaluations of conduct problems.

In the points which follow, the conclusions are restricted usually to parent's evaluations of the children. Also, it should be kept in mind that even though the conclusions are based mostly on cross-parental ratings to avoid inflation due to rater bias, one cannot assume that all judgmental bias has been eliminated in the evaluation of child behavior.

4. On the basis of previous work it was expected that the number of associations between father and child behavior would equal or exceed that for mothers. This hypothesis was confirmed. Again using only uncontaminated correlations 22% of the father-in-relationship-to-child correlations were significant, while only 9% of the corresponding mother correlations were significant. A possible artifactual explanation for the marked difference was considered.

5. Father's ratings of his ideal relationship to his child produced a number of significant correlations with child behavior. If father's

conception of his ideal relationship is more loving, democratic, and emotionally mature, the child is rated by mother as better adjusted, outgoing, less demanding, etc. Interestingly, the father ideal ratings added new variance in the prediction of child behavior from parent behavior.

6. The hypothesis that child Conduct Problems are related to general maladjustment, open venting of negative emotions, and arbitrary and inconsistent discipline on the part of both parents was given considerable support.

7. The hypothesis that child Personality Problems are related to paternal maladjustment and autocracy and are independent of maternal behavior was only partly supported. Maladjustment of the father, characterized by hostility and general social withdrawal, was found to be associated with child Personality Problems. Associations with mother's behavior depended upon who was rating whom.

8. Discrepancies between factor scores (D^2) for mother and father ratings of father were found to be significantly related to child Conduct Problems.

9. Analysis of second-order evaluative factors among the parent factor scores revealed a number of significant relationships to child problem behavior. There is a strong indication that the general positiveness or negativity of parental attitudes is critical in child adjustment.

10. As in the previous study (Becker, et al., 1959), the findings point emphatically to the importance of the need for more systematic study of the role of the father in child development.

REFERENCES

- BALDWIN, A. L., KALHORN, JOAN, & BREESE, FAY H. The appraisal of parent behavior. *Psychol. Monogr.* 1949, 63(4, Whole No. 299).
- BANDURA, A., & WALTERS, R. H. *Adolescent aggression*. New York: Ronald, 1959.
- BECKER, W. C., PETERSON, D. R., HELLMER, L. A., SHOEMAKER, D. J., & QUAY, H. C. Factors in parental behavior and personality as related to problem behavior in children. *J. consult. Psychol.*, 1959, 23, 107-118.
- CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book, 1957.
- CATTELL, R. B., & COAN, R. W. Child personality structure as revealed in teachers' behavior ratings. *J. clin. Psychol.*, 1957, 13, 315-327. (a)

- CATTELL, R. B., & COAN, R. W. Personality factors in middle childhood as revealed in parents' ratings. *Child Develpm.*, 1957, 28, 439-458. (b)
- DIAMOND, S. *Personality and temperament*. New York: Harper, 1957.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- HEWITT, L. E., & JENKINS, R. L. *Fundamental patterns of maladjustment: The dynamics of their origin*. Springfield, Ill.: State of Illinois, 1946.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.
- PETERSON, D. R., BECKER, W. C., HELLMER, L. A., SHOEMAKER, D. J., & QUAY, H. C. Parental attitudes and child adjustment. *Child Develpm.*, 1959, 30, 119-130.
- PETERSON, D. R., & CATTELL, R. B. Personality factors in nursery school children as derived from parent ratings. *J. clin. Psychol.*, 1958, 14, 346-355.
- ROFF, M. A factorial study of the Fels Parent Behavior Scales. *Child Develpm.*, 1949, 20, 29-45.
- SEARS, R. R., MACCOBY, ELEANOR E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.
- SILVERSTEIN, JOAN. A study of relationships within the families of underachievers and overachievers. Unpublished bachelor's thesis, University of Illinois, 1958.
- WHITING, J. W. M., & CHILD, I. L. *Child training and personality*. New Haven: Yale Univer. Press, 1953.

(Received November 25, 1959)

EXPERIMENTAL MODIFICATION OF INTERVIEWER CONTENT IN STANDARDIZED INTERVIEWS¹

FREDERICK H. KANFER

Purdue University

JEANNE S. PHILLIPS, JOSEPH D. MATARAZZO, AND GEORGE SASLOW

University of Oregon Medical School

In contrast to research on psychotherapy which studies outcome, or the substantive content of interviews and the patient's reported life experiences outside of the psychotherapy session, a number of recent approaches have focused upon the immediate interactional behavior of therapist and patient within the interview itself. The descriptive dimensions within which the interactional behaviors have been classified are multiple and varied, ranging, for example, from Snyder's (1945) early study of the effects of assumption of responsibility by the therapist on the patient's release of feelings, expressions of problems, insight, etc. to quantitative analyses of temporal characteristics of the interview interactions (Saslow & Matarazzo, 1959).

In addition to contributing to a better understanding of the interpersonal processes of psychotherapy and to personality theory, these examinations of relationships between various characteristics of the verbal behavior of the two participants have shed light upon specific effects of the interviewer's actions upon the subject's (S's) subsequent behavior during the session. Recent studies, for example, have demonstrated that patient-categories "approved" by the therapist increase while

those "disapproved" decrease throughout the course of therapy (Murray, 1956), and that use of subtle verbal reinforcers by the interviewer can significantly increase the frequency of affect responses in a group of schizophrenics (Salzinger & Pisoni, 1958). Numerous authors (see the summary by Lacey, 1959) have found changes in patient's autonomic responsiveness to be related to the therapist's behavior. In a series of studies utilizing a standardized interview technique first described by Chapple (1953), Saslow, Goodrich, and Stein (1956), Saslow, Matarazzo, and Guze (1955), Saslow and Matarazzo (1959), and Matarazzo, Saslow, Matarazzo, and Phillips (1958) have repeatedly demonstrated an interviewer effect on the temporal aspects of the patient's utterances when the interviewer's own verbal behavior is varied along the same time-unit parameters. Goldman-Eisler (1952) also reports this effect in a somewhat less restricted setting. Similarly, Kanfer (1959) found variations in verbal rate when interview topics and experimenter's (E's) participation in the interaction were controlled.

When interviews are conducted with therapeutic intent, interpretations have long been considered the most crucial to successful outcome. In several psychoanalytically oriented theories, it is held that the major changes in a patient's behavior result from interpretations which make conscious those unconscious motivational factors which determine the patient's actions. Rogers (1942), on the other hand, states that insight and consequent behavioral change is "often delayed, and sometimes made impossible, by efforts of the coun-

¹ This study was supported in part by Research Grants M-1938 and M-2027(C1) from the National Institute of Mental Health, of the National Institutes of Health, United States Public Health Service. The data were collected while the first author was Visiting Professor in the University of Oregon Medical School during the summer of 1958.

We wish to acknowledge the assistance of Bernadene Allen and Thomas Cunningham in the analysis of the data.

selor to create it or bring it about" (p. 195). While dispute thus exists as to the direction or desirability of the effects of interpretations, there certainly is none as to the presence of significant effects, or the need to assess carefully their locus and quality. Consistent with interest in immediate interactional consequences of different classes of therapist actions, interpretations may be viewed as having an impact on the further course of the psychotherapeutic process not only as a function of their substantive qualities or veracity, but also through changing the patient's current interview interactional pattern (cf. Auld & White, 1959; Speisman, 1959).

Especially in its early stages the therapeutic task is characteristically defined as maintaining a flow of communication between patient and therapist. Therefore, the temporal dimension (duration) of the patient's verbal output at any moment has repeatedly been considered as an important measure of his response to various therapist operations. Since a direct test of the effect of experimentally controlled interpretations in psychotherapy presents many methodological and practical problems, one possible approach is to study the influence of interpretations during standardized personal interviews, modeled after the clinical interaction in which they are usually given. While several alternative consequences of interpretations for the duration of the interviewee's subsequent utterances can be considered and further refinement of the definition of interpretations as a class is also feasible, the *basic hypothesis of the present study is simply that interpretations, as a global category of interviewer's behaviors, differ from exploratory or information seeking statements in their immediate effect on the interviewee's verbal output, in time units, regardless of their differing content or their role in the interview strategy.* This implies that all interpretations share to some extent a broad similarity with respect to the interviewee's perception of their significance in the immediate interpersonal relationship. Hence responses to them as a distinct class of interviewer actions would also be similar.

The direction of the change in the interviewee's duration of utterances may also be a

function of several other factors which determine the interviewee's perception of the interpretation, such as the length of familiarity with the interviewer, the particular phase of psychotherapy, the degree of elaboration with which it is given, the various characteristics of the patient, etc. By using a single initial interview in this study some of these variables were brought under strict experimental control, albeit at the expense of sacrificing some of the richness inherent in the usual therapy sessions.

As has been mentioned, extensive work of several groups of investigators on normals and patients with measures of verbal behavior in time units has already clearly demonstrated the existence of functional relationships between such variables as duration of speech and duration of silence on the part of interviewer and interviewee. Indeed, it was the recognition of such relationships which led Chapple (1953) to develop a partially standardized interview in order to eliminate contamination and unreliability in the use of the interview as a research tool. This technique utilizes temporal variables, recorded by the Interaction Chronograph (Chapple, 1949), for which strikingly high reliability has been established both across different interviews and over time, for a number of patient and normal samples (summarized in Saslow & Matarazzo, 1959). Relationships have also been found between these time-unit characteristics of the interviewee's interaction pattern and a variety of organismic variables (Matarazzo, Matarazzo, Saslow, & Phillips, 1958). The present study permits exploratory investigation of possible relationships between these *temporal* variables and *functional content* within the same interview. Specifically it asks whether changing *what* the interviewer says (interpretation) while controlling *how* he says it (time-unit variables) has any effect upon the stable *temporal* aspects of the S's utterances. The latter are already known to be responsive to modification of *how* (in time units) the interviewer speaks.

METHOD

Subjects

The 60 Ss in this study were female volunteers drawn from the hospital staff and student nurse

populations of the University of Oregon Medical School.²

Procedure

Ss were told that they would be seen for a personal interview, conducted by a visiting research psychologist who was not on the staff of the school, and that the interview would be confidential and had no relationship whatever to their status as students or nurses, respectively.

The Ss were asked to talk freely about any topic and were told that the interview was designed to find out how to best conduct interviews with psychiatric patients whose problems need to be discussed in order to help them get well. The Ss were asked if they understood the procedure. From their comments at this time and from later postexperimental interviews it was clear that the Ss perceived the interview as a discussion of their personality and their emotional attitudes and problems and that they had expected to gain some insight into their own behavior. Ss were again assured of confidentiality and the interview began with the question: can you tell me how you first became interested in nursing?

Ss from three samples which differed with respect to age and professional status (student nurses vs. hospital medical vs. psychiatric staff nurses) were randomly assigned to experimental and control groups. All interviews were conducted individually by the same *E*, in an office with a one-way screen. The time-unit variables of the interaction were recorded on the Interaction Chronograph by the same experienced observer who also tape recorded each session. Although the Interaction Chronograph yields several different variables, the present study concerned itself only with Action, the average duration of the verbal action of *S* and *E*.

The interview derived from one described by J. D. Matarazzo et al. (1958) and consisted of three periods: Period I lasted approximately 15 minutes, while Periods II and III lasted 10 minutes each. For all periods, the rules for *E*'s behavior restricted only the duration of *E*'s utterances (5 seconds), and required that *E* reply within 1 second after *S* ceased talking. In those rare instances where *S* failed to reply, *E* talked again 15 seconds later. All verbal, gestural, and postural behavior on the part of *E* was limited to the 5-second interval as described above. Several practice interviews, not used in the study, were conducted by *E* under observation.

In this study *E*'s behavior was equivalent in Periods I and III in all groups. *E*'s comments consisted of exploratory, information getting questions. *E* controlled the content and approximate sequence of topics by introducing questions in the following areas: vocation, family, friends, social and recrea-

tional activities. The same topics were discussed in each of the three interview periods for all groups.

The major independent variable was the modification of the content of *E*'s utterances in Period II. For all control groups this period was identical in all respects to Periods I and III. For all experimental groups, *E* introduced the period by an information seeking question. He then followed with 12 interpretive statements, separated by as many noninterpretive (information seeking) comments as needed to maintain the natural tone of the interview. For the purpose of this study, interpretations were defined as any statements, often phrased as questions, which: analyze; relate several of *S*'s described experiences, attitudes, or feelings; generalize *S*'s statements to suggest a determinant of his behavior; or give a psychological explanation; or a motivational description. All intended interpretive statements were signaled in advance by *E* to the observer and recorded. No interpretive statement referred directly to *S*'s motor behavior in the interview or to his relationship with *E*. The topics in this period were the same as in Periods I and III and no new topic was introduced by *E*. The following examples, randomly selected from the interview transcriptions, illustrate the two classes of statements.

Exploratory (noninterpretive) statements:

1. I wonder if you can tell me a little bit more about what kind of things you enjoy doing when you are off duty.
2. You have told me about your family. Now I wonder if you can tell me a little about your father, what kind of person is he?
3. When you decided to go into nursing, what kind of satisfactions did you think you were going to get out of your work as a nurse?
4. Could you tell me a little about the kind of things that you came across in your work which seem irritating to you?

Interpretive Statements:

1. It seems that you're the kind of person who avoids hurting other people, sometimes even at the expense of hurting yourself.
2. When you see a patient get better it makes up for some of the suffering that you yourself had to bear.
3. Doing a conscientious and thorough job makes you feel better, and you feel that you are an important person.
4. I imagine that you resented the fact that sometimes when you are with him you almost feel like a little child yourself.

In Period III *E* again limited his actions to exploratory, information getting questions. At the end of the interview, *E* discontinued his standardized role and asked several specific questions concerning *S*'s feelings about the interview in an informal manner.

Since it was expected from previous findings that change in Action in response to the interpretation variable might be related to the level of maturity

² We wish to thank the Departments of Nursing Education and Nursing Service and their Directors, Jean Boyle and Shirley Thompson, of the University of Oregon Medical School, for their valuable assistance in obtaining Ss.

and the extent of Ss' life experiences, as well as familiarity with psychological procedures, the additional variables of age and professional status were introduced by replication of the procedure with student nurses, and nurses who were working on the Neuropsychiatric ward. The design resulted in the following five groups:

1. Student nurses, experimental (interpretations given) (SN-E), $N = 15$.
2. Student nurses, control (no interpretations given) (SN-C), $N = 10$.
3. Graduate nurses, medical staff, experimental (GN-E), $N = 15$.
4. Graduate nurses, medical staff, control (GN-C), $N = 10$.
5. Graduate nurses, NP staff, experimental (NP-E), $N = 10$.

With this design the control groups totaled 20 Ss, the experimental groups totaled 40 Ss.

RESULTS

The mean durations of Ss' utterances for all groups over the three periods are shown in Table 1. Inspection of the table shows a general trend of all experimental groups to decrease in mean Action during Period II and to recover in Period III. These data were subjected to an analysis of variance, using an unweighted means analysis to take into account unequal frequencies in the subclasses. The summary of the analysis is presented in Table 2. As indicated by the nonsignificant F between experimental and control groups

TABLE 1
PERIOD MEANS AND STANDARD DEVIATIONS
FOR Ss' ACTIONS^a

Group		Period			F
		I	II	III	
NP-E ($N = 10$)	Mean	43	31	44	4.50*
	SD	33	15	21	
GN-E ($N = 15$)	Mean	44	34	59	22.38**
	SD	23	15	27	
SN-E ($N = 15$)	Mean	33	23	36	6.70**
	SD	12	8	14	
GN-C ($N = 10$)	Mean	38	37	39	0.07
	SD	26	19	29	
SN-C ($N = 10$)	Mean	44	50	48	0.95
	SD	22	16	27	

^a Measured in seconds.

* $p < .05$.

** $p < .01$.

TABLE 2
ANALYSIS OF VARIANCE OF MEAN ACTIONS

Source	df	MS	F
Between Ss	59		
Exper. vs. Control (A) ^a	1	152.01	0.58
Student vs. Staff (B)	1	70.25	0.27
$A \times B$	1	1,245.24	4.75*
GN-E vs. NP-E	1	174.03	0.66
Ss within groups	55	261.96	
Within Ss	120		
Periods (P)	2	258.14	9.78**
$A \times P^b$	2	279.02	10.57**
$B \times P$	2	44.65	1.69
$A \times B \times P$	2	39.42	1.49
GN-E vs. NP-E $\times P$	2	84.63	3.21*
Ss in groups/periods	110	26.40	
Total	179		

^a The partial breakdown of the between-groups sum of squares, with 4 df , compares SN-E and GS-E vs. SN-C and GS-C (A); SN-E and SN-C vs. GS-E and GS-C (B); the AB interaction; and an a priori comparison of GN-E vs. NP-E which is nonorthogonal to the first three.

^b The group \times period interaction, with 8 df , was broken down into interactions derived from the student and general staff groups, with (GN-E vs. NP-E) $\times P$ representing a separate comparison of special interest to our hypotheses.

* $p < .05$.

** $p < .01$.

(A), the duration of utterances for the total interview did not differ for the groups. However, mean Action differed significantly over the three periods ($F = 9.78$). In addition, periods interacted significantly with the treatment variable A. Therefore, tests for the simple effects of periods for each group were run (see Table 1). They show that the significant treatment \times periods interaction ($F = 10.57$) is due to the effects of the experimental treatment. The experimental groups show significant changes over periods (for NP-E, $F = 4.50$; for GN-E, $F = 22.38$; and for SN-E, $F = 6.70$), while the control groups show no change in mean Action over the periods (for GN-C, $F = 0.07$; for SN-C, $F = 0.95$).

The effect of interpretations during Period II is seen by a decrease in the average length of utterances immediately following the interpretive responses. However, when the interpretations are discontinued in Period III, the mean duration of utterances increases considerably. The pattern of decreased Action in Period II and increased Action in Period III is shown by 31 of the 40 experimental Ss and by 4 out of 20 control Ss. A chi square test

for these frequency distributions is significant beyond the .01 level. In this "aftereffect" of the interpretation phase, the psychiatrically less sophisticated Ss in Group GN-E differ significantly from the NP-E group ($F = 3.21$) in increasing the duration of their verbalizations beyond their mean for Period I (see Table 1). The student nurses similarly exceed their mean Period I Actions subsequent to interpretations. The NP-E group, however, returns to their Period I mean.

The comparisons between student and staff nurses yield a significant F ratio when the interaction between treatment (A) and age (B) is considered (Table 2). A test for the simple effects of the age factor in each treatment yielded a significant F ratio for the comparison of SN-E vs. GN-E, but a comparison of Groups SN-C and GN-C showed no significant differences. Therefore, the differences in age and experience between the student and staff nurses led to a different response only when interpretations were given in the interview. From the group means (Table 1) it is seen that the student nurses showed a lesser tendency to increase in mean Action during Period III than did the GN-E group.

In order to further evaluate these findings, two possible sources of bias were examined. One basic assumption of the present study is that E 's statements can be separated into a class of information seeking responses and a class of interpretive responses. The validation of this assumption was therefore necessary in order to meaningfully describe the independent variable. Fifty-five of E 's statements were randomly selected to represent all Ss and all three periods. This sample was given to 12 judges, all clinical psychologists with at least 2 years' postdoctoral experience, who were asked to order the statements into two categories, i.e., interpretive and noninterpretive statements, according to the definition given earlier. Judges' agreements were compared with E 's signaled intent during the experiment.

On 48 out of the 55 items there was 100% agreement among the 12 judges in ordering the statements as they were used and intended by E . At least 10 of the 12 judges agreed with E 's usage on five more of the statements. On the two remaining statements, 6 judges agreed

with E 's use of the statement as representative of an interpretation or noninterpretation. Thus, there was unanimous agreement by the judges with E 's usage on 87.3% of all sampled statements while only 3.6% of the sampled statements presented any serious difficulties in categorizing them. These results indicate that interpretive statements (the independent variable) and the control statements are clearly discriminable as separate classes and were applied in accordance with the prescribed procedure.

A second assumption is that E , himself, did indeed conform to the rules governing his utterances in the standardized interview and that his behavior in the experimental and control interviews was therefore essentially identical except for the presence or absence of the independent variable, the interpretations in Period II. To ascertain whether this assumption had been met, several analyses were performed. First, to test the assumption that the durations of the three periods and of the total interview were as prescribed (15', 10', 10', and 35', respectively), their mean durations were compared for all groups. The ranges of mean duration were: Period I, 15.47 to 16.09 minutes; Period II, 10.10 to 12.05 minutes; Period III, 9.20 to 10.47 minutes; and the length of the total interview, 35.76 to 37.85 minutes. The very small differences in period duration were randomly distributed over the groups. These results therefore indicate that the prescribed control of period length by E had been fulfilled.

Another primary aspect of the standardizing rules has to do with duration of E 's utterances (E 's Action), i.e., that E 's utterances average approximately 5 seconds in length. The duration of individual utterances by E was obtained and averaged separately for each period for each S. All of E 's Actions in Period II were utilized, while 10 or more utterances in Periods I and III were randomly sampled for each S, after this sampling had been demonstrated to be a highly reliable and valid measure of the durations of all utterances within these periods. The means and SDs of E 's Actions are shown in Table 3. F tests were then computed for mean E 's Action across the three periods, for each group. While the values for F indicate significant dif-

TABLE 3
E's ACTION: PERIOD MEANS^a AND
 STANDARD DEVIATIONS

Group		Period		
		I	II	III
NP-E (<i>N</i> = 10)	Mean	5.46	6.98	5.30
	<i>SD</i>	.82	.71	.78
GN-E (<i>N</i> = 15)	Mean	6.53	8.52	6.73
	<i>SD</i>	.93	1.56	1.18
SN-E (<i>N</i> = 15)	Mean	5.21	7.21	5.74
	<i>SD</i>	.71	1.35	.75
GN-C (<i>N</i> = 10)	Mean	6.25	7.19	7.89
	<i>SD</i>	.77	1.14	1.75
SN-C (<i>N</i> = 10)	Mean	6.38	6.90	6.71
	<i>SD</i>	.45	.88	.88

^a Measured in seconds.

ferences ($p < .01$) in mean *E*'s Action between periods for all groups except SN-C, the absolute difference between successive period means was never greater than 2.00 seconds, while 4 of the 10 differences were less than 1 second. The very small magnitude of *E*'s variation in length of utterance thus suggested that the statistically significant differences between periods could not be responsible for any differences in *S*'s Action between periods. As can be seen in Table 1, the differences between means obtained for *S*'s Action were always at least five times as great as those for *E* and usually this factor was much larger.

However, in view of the significant values for *F* for *E*'s Action across periods, two correlational analyses were carried out to investigate whether these significant but small variations in *E*'s Action had had any biasing effects on *S*'s Action. Rho's were computed by groups between each *S*'s mean Action and *E*'s concurrent mean Action for each period. Means for *E* and *S* in Period II were separated into two parts, interpretations and noninterpretations. Out of the 18 rho's computed, only 3 were significant. For SN-E, the rho for Period II (noninterpretations) was .54, $p < .05$; and for Period III, rho was .66; $p < .05$. The other significant correlation, for Period IIb (interpretations) in the NP-E group, was -.84, $p < .01$. Examination of *SD*s, sample

size of *S*s, their number of utterances per period, and other data did not help in interpreting these quite inconsistent findings. The absence of any correlation for 15 out of 18 comparisons suggested that there was no systematic relationship between *E*'s Actions and *S*'s Actions because *E*'s Actions were relatively constant. This conclusion was supported when correlations were obtained *individually* for each of the 25 *S*s in the groups for whom the rho's above were significant (SN-E and NP-E). That is, rho's were computed using *each single* utterance by *E* and the *following* single utterance by *S* as the scores. None of these 25 correlations was significant.

Additional evidence supporting the conclusion that the minor variation in *E*'s Action, while undesirable, was not a contaminating factor in the experimental results, is the fact that previous studies have shown a *positive* correlation between the Actions of two interacting individuals when their Actions are left free to vary (Saslow & Matarazzo, 1959). Therefore, if any bias due to lack of constancy in *E*'s Actions had occurred in the present study, it would be expected to *lengthen* *S*'s utterance after *E*'s *longer* interpretations, whereas in fact, the *opposite* changes were found in *S*s' Actions in Period II.

DISCUSSION

The results of the present study clearly indicate that interpretations affect the temporal duration of *S*'s utterances, both immediately and subsequently in a standardized interview which focused on personal material. Therefore, the findings suggest that interpretations are a class of interviewer responses which have a distinct consequence on the interviewee's behavior and thereby modify the subsequent interaction. The data thus support the hypothesis that a class of interpretive responses leads to differing interactional behavior on the part of *S* than exploratory, information seeking questions.

In extending the gross relationships suggested by the present findings, it would be appropriate to ask further questions concerning the *specific* properties in which members of the general class of interpretive responses differ. For example, can more highly differ-

entiated effects on the immediate and subsequent temporal interaction pattern of *S* be attained by systematic variation of the location of interpretive statements on a dimension of "depth," appropriate timing, theoretical origin, etc.? That any member of a class of interpretations contributes toward modification of the temporal interaction pattern of an interviewee by no means reduces the likelihood of additional effects attributable to further subdimensions within the broad class of interpretive statements.

When making interpretations in psychotherapy there is usually a shift in the therapist's role from permissive listener to a person who evaluates and "knows" the patient and corrects his self-view by presenting for consideration alternate explanations. Thus, at least those interpretations which "uncover unconscious motives" might be considered as punishing stimuli (Murray, 1954). Dollard and Miller (1950) also imply the aversive consequences of interpretations. The reduction in verbal output demonstrated in this study may then relate to a punishing aspect of the interpretation. An alternate possible explanation of the present data is that interpretations differ from exploratory statements simply because they usually contain clear-cut statements and can more easily be answered by a simple yes or no. Although the phrasing of interpretations often results in a *question*, starting with "do you think that . . ." or "wouldn't you say . . .," these questions may be rhetorical and ask simply for direct agreement or disagreement with the therapist's statement. However, they should certainly provoke more discussion, qualification, or even rebuttal than a single word. In the present study, none of the *Ss* actually replied with such brief assent or disagreement. If interpretations, as a class, intrinsically lead to more brief responses by *S*, then this property of interpretations alone is well worth considering. If interpretations had led to very brief utterances ("yes" or "no") only in immediate response to the interpretive statement, then *Ss* in the experimental groups during Period II might be expected to show a much larger range and standard deviation from their mean, since both interpretive and noninterpretive statements were given in Pe-

riod II. None of the groups showed such increased variability or range. In fact, the distributions for Period II, including both interpretive and noninterpretive statements, were somewhat narrower in the experimental groups than in their appropriate control groups.

The specific content and dynamic importance of an interpretation probably also have some long term effects which are neither apparent nor measurable during the interaction immediately succeeding it. In our postexperimental questioning of *Ss*, almost all of the *Ss* in the experimental groups mentioned an interpretation when asked to recall that statement by *E* which had stood out most in the interview. By contrast, the control *Ss* did not recall any specific questions which the interviewer had asked. *Ss* in the experimental groups tended to discuss again the interpretation itself and their attitudes about it, indicating that *E*'s interpretive comments had stimulated self-exploration, beyond the limits of the interview itself. Further research would be needed to evaluate the effects of these interpretations on subsequent sessions.

The present study, like most current research on clinical phenomena, is an attempt at defining the effect of one group of variables by means of measuring behavioral changes which themselves have not yet been fully correlated with central personality characteristics. In this sense it represents an additional one in a series of studies aimed at exploring a standardized interview and investigating its utility for the examination of clinical hypotheses by means of objective recording of two-person interactions. The results obtained here add to the evidence that the temporal interaction measures have validity not only with respect to a variety of temporal stimulus variables which modify *E*'s temporal behavior pattern (Saslow & Matarazzo, 1959), to organismic variables (R. G. Matarazzo et al., 1958), and to some aspects of verbal content (Phillips, 1957), but also with respect to the functional dimension of verbal stimulus variables. That is, these variables appear sensitive to changes in the interpersonal role of *E*'s utterances when his temporal patterns are under experimental control. Further, the finding that the younger student nurses responded to interpretations

with shorter utterances may be thought of as consistent with the earlier finding (R. G. Matarazzo et al., 1958) that younger persons seem to respond to potentially aversive stimuli of the standardized interview (e.g., *E*'s silence) with reduced verbal output. It would seem then, that there is a partially equivalent response in the younger, less experienced, or less "poised" person to controlled changes in *either* the functional content *or* certain time-unit characteristics of *E*'s behavior, even though *E*'s behavior would seem to be more assertive when he makes interpretations and more passive when he fails to respond and remains silent.

Another implication of the present results is worthy of mention. This concerns the broad question of the effect of the *content* of the interview on interviewee interaction behavior. In 1940 Chapple made the interesting suggestion that the actual *content* discussed by a person would affect the other person's reaction to him much less than would the *timing* of the first person's interactions (Chapple & Arensberg, 1940, p. 33). The results of the present study, while not dealing specifically with effects upon the content of the interviewee, but, in fact, with changes in the interviewer's gross *content* (i.e., from information getting content to interpretive content), suggests, at least, that interviewer induced changes in content *do produce* a demonstrable change in the *interaction behavior* of the interviewee. We are currently designing a study in which predefined and different content areas will be discussed with a given *S*, thus providing a more direct test of Chapple's suggestion that interviewee content differences may correlate only minimally with interviewee interaction patterns.

SUMMARY

This study investigated the hypothesis that interpretations as a category of interviewer responses differ from exploratory or information seeking statements both in their immediate and their subsequent effects on the duration of interviewees' utterances, regardless of their specific content or their role in the interview strategy, at least in standardized experimental interviews.

Sixty female nurses were interviewed by

means of a standardized interview which limits *E*'s verbal and gestural participation in the interaction to 5-second durations. The interview consisted of three periods. During Periods I and III *E* made only exploratory or information seeking statements for all *Ss*. During Period II *E* gave 12 interpretations to *Ss* in the experimental groups, while continuing exploratory statements for the control groups. The procedure was replicated with student nurses, medical nurses, and psychiatric nurses since age and experience were expected to affect *S*'s response to the interpretations. All interactions were recorded by the Interaction Chronograph and the sessions were tape recorded.

The results showed significantly *lower* mean durations of utterances for the experimental groups during Period II (the interpretation period). The control groups showed no change in mean duration of utterances over the three periods. Furthermore, a significantly *higher* mean duration of verbalizations was found in Period III in these groups which had been given interpretations in the preceding period. Student nurses did not differ from older staff nurses in the control groups, but yielded a significantly lower verbal output when interpretations were given.

The results showed that interpretations as a class of interviewer techniques have a distinct effect on interviewee's interactional patterns in the standardized interviews used in this study, with normal *Ss*. Interpretations tend to shorten the duration of utterances immediately following them. In addition, after interpretations cease, interviewees tend to increase the duration of their verbal output in comparison with *Ss* who were not given any interpretations at all. The interviewees' age, experience, or both tend to further modify their reactions to interpretations.

REFERENCES

- AULD, F., & WHITE, ALICE M. Sequential dependencies in psychotherapy. *J. abnorm. soc. Psychol.*, 1959, 58, 100-104.
- CHAPPLE, E. D. The Interaction Chronograph: Its evolution and present application. *Personnel*, 1949, 25, 295-307.
- CHAPPLE, E. D. The standard experimental (stress) interview as used in Interaction Chronograph investigation. *Hum. Organization*, 1953, 12, 23-32.

- CHAPPLE, E. D., & ARENSBERG, C. M. Measuring human relations: An introduction to the study of the interaction of individuals. *Genet. psychol. Monogr.*, 1940, 22, 3-147.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- GOLDMAN-EISLER, FRIEDA. Individual differences between interviewers and their effect on interviewees' conversational behavior. *J. ment. Sci.*, 1952, 98, 660-671.
- KANFER, F. H. Verbal rate, content, and adjustment ratings in experimentally structured interviews. *J. abnorm. soc. Psychol.*, 1959, 58, 305-311.
- LACEY, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: APA, 1959. Pp. 160-208.
- MATARAZZO, J. D., SASLOW, G., MATARAZZO, RUTH G., & PHILLIPS, JEANNE S. Stability and modifiability of personality patterns during a standardized interview. In P. A. Hoch & J. Zubin (Eds.), *Psychopathology of communication*. New York: Grune & Stratton, 1958. Pp. 98-125.
- MATARAZZO, RUTH G., MATARAZZO, J. D., SASLOW, G., & PHILLIPS, JEANNE S. Psychological test and organismic correlates of interview interaction behavior. *J. abnorm. soc. Psychol.*, 1958, 56, 329-338.
- MURRAY, E. J. A case study in a behavioral analysis of psychotherapy. *J. abnorm. soc. Psychol.*, 1954, 49, 305-310.
- MURRAY, E. J. A content-analysis method for studying psychotherapy. *Psychol. Monogr.*, 1956, 70(13, Whole No. 420).
- PHILLIPS, JEANNE S. The relationship between two features of interview behavior comparing verbal content and verbal temporal patterns of interaction. Unpublished doctoral dissertation, Washington University, St. Louis, 1957.
- ROGERS, C. *Counseling and psychotherapy*. Cambridge: Houghton Mifflin, 1942.
- SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, 57, 84-90.
- SASLOW, G., GOODRICH, D. W., & STEIN, M. Study of therapist behavior in diagnostic interviews by means of the Interaction Chronograph. *J. clin. Psychol.*, 1956, 12, 133-139.
- SASLOW, G., & MATARAZZO, J. D. A technique for studying changes in interview behavior. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: APA, 1959. Pp. 125-159.
- SASLOW, G., MATARAZZO, J. D., & GUZE, S. B. The stability of Interaction Chronograph patterns in psychiatric interviews. *J. consult. Psychol.*, 1955, 19, 417-430.
- SNYDER, W. U. An investigation of the nature of non-directive psychotherapy. *J. gen. Psychol.*, 1945, 33, 193-223.
- SPEISMAN, J. C. Depth of interpretation and verbal resistance in psychotherapy. *J. consult. Psychol.*, 1959, 23, 93-99.

(Received December 1, 1959)

DEVELOPMENTAL ASPECTS OF IMPULSE CONTROL¹

NATHANIEL N. WAGNER

Astor Home for Children, Rhinebeck, New York

The concepts impulsivity and impulse control appear frequently in the vocabulary of the clinician who works with children. Such writers as Redl and Wineman (1957) attach extreme importance to the "control system" in the understanding and treatment of emotional disorders in children. Aside from Goldfarb's (1949) and Colvin's (1958) work on the effects of institutionalization upon impulse control, this writer has been unable to find any experimental investigation on the development of impulse control. This condition is all the more remarkable in view of the long history of theoretical interest in impulsivity (Freud, 1938; Rapaport, 1951; Schilder, 1951) and the many studies on the relation of Rorschach movement and other measures of fantasy to motor inhibition (Meltzoff, Singer, & Korchin, 1953; Singer & Herman, 1954; Singer, Wilensky, & McCraven, 1956).

Colvin (1958) studied the relation between parental deprivation, in the form of early and prolonged child care institutionalization, and various measures of ego and social development. His results indicated a significant relationship between the extent of institutionalization and impulse control with the lack of opportunity for parental identifications seen as a major factor in the low impulse control of the institutionalized child. The present investigation focuses upon other characteristics of the child's interpersonal experience, i.e., the effects of specific aspects of the nature and quality of the parent-child interaction upon the development of impulse control.

¹ Presented at the APA meetings at Cincinnati, September 7, 1959, in a symposium, Manifestations and Measures of Impulse Control and Its Behavioral Concomitants.

The author wishes to thank Ralph W. Colvin for his assistance in all phases of this study.

PROCEDURE

The 36 subjects were the entire population of the Astor Home for Children, a residential treatment and research center for emotionally disturbed boys. Ages ranged from 7 years 1 month to 15 years 3 months, with a mean of 11 years; WISC IQs from 68 to 119, with a mean of 97; diagnoses from schizophrenia to primary behavior disorder; and symptoms from severe withdrawal to explosive acting out.

To secure a measure of impulse control, behavior rankings were made independently by five professional staff members, all of whom could rank the entire population. Each judge was instructed to rank the 36 students on impulse control which was defined as: "ability to delay immediate satisfactions for long range goals . . . , a capacity to restrain action until the formulation of a plan" (Colvin, 1958). Intercorrelations among judges ranged from .77 to .92, with a coefficient of concordance of .84, which is significant beyond the .001 level. Therefore, the ranking of the sum of the ranks was used as the "best" ranking (Walker & Lev, 1953).

These exceptionally high intercorrelations certainly, to some extent, reflect the high level of communication between staff members, which is an integral part of residential treatment. However, all rankers were instructed not to communicate with each other concerning these specific impulse control rankings. It is this writer's impression that all five staff members fulfilled this condition. None the less, the probability of earlier staff meetings and other types of communications influencing the level of ranker agreement cannot be ruled out.

The Bloch and Behrens Multiproblem Family Index (Bloch, Behrens, Guttenberg, King, & Tendler, 1959) was selected as the measure of parent-child interaction to be related to the development of impulse control. The following six areas comprise the index:²

1. Marriage: Either parent abandoned child as infant or after one year; there was a marital history of separation and reunion; the marriage was broken (other than by death).

2. Parent-child relationship: Either parent was abusive toward child, beat him, over-disciplined or was punitive; the atmosphere in the home was grossly psycho-pathogenic.

² Quoted by permission of the authors.

3. Social pathology: Diagnosis of specific social pathology in any family member, such as alcoholism, drug addiction; either parent or a sibling was in a penal institution at time of child's application to a residential treatment center.

4. Physical illness: Either parent had a chronic physical illness with or without some incapacity; either parent had a serious physical illness which interfered with his parental and family functioning.

5. Mental illness: Either parent or a sibling was hospitalized for mental illness; either parent, sibling, or other family member had a psychiatric diagnosis of mental illness (whether psychotic or non-psychotic) or mental deficiency.

6. Financial support: Family was either wholly or partly dependent on public assistance funds at time of child's application to a residential treatment center or in the past, family housing was grossly inadequate.

The incidence of these areas of family pathology was computed for each child through an interview with two professional child care workers thoroughly familiar with the children's case histories. Where there was disagreement, or where either of the workers expressed some doubt, the experimenter consulted the case history material and made the decision. It was necessary to consult the case histories in six cases: four as a result of disagreement, two where there was some doubt expressed. The degree of agreement ranged from 92% to 100% on the individual areas and was 97% for the total index scores.

The scorings on the index were kept independent from the rankings of impulse control. The rankings had been obtained in the context of another study conducted 5 months previous to the scoring of the

Multiproblem Family Index. Also, the staff members who acted as informants on the index were not aware that it would be related to impulse control.

The population was divided into high and low impulse control groups through a division at the median of the rankings with 18 boys in each group. Family pathology, as measured by the Multiproblem Family Index, was then related to the impulse control level of the subjects by means of chi square tests of the relationships between: (a) high or low impulse control and the absence or presence of each area of pathology, (b) high or low impulse control and the low (0-3) or high (4-6) incidences of the six areas of family pathology.

RESULTS

Table 1 presents the two-by-two contingency tables relating impulse control with each of the areas of the Multiproblem Family Index. All of the areas, with the exception of physical illness, are significantly related to impulse control in that their presence is, associated with low impulse control, their absence with high impulse control.

Table 2 presents the relationship of high and low impulse control and the extent to which the subjects were exposed to the various areas of family pathology. It indicates that children with high impulse control tend to come from families having no or few problem areas present, whereas the low impulse control group tends to come from families

TABLE 1
COMPARISON OF 18 HIGH AND 18 LOW IMPULSE CONTROL SUBJECTS ON EACH
AREA OF THE MULTIPROBLEM FAMILY INDEX

Multiproblem Family Index Areas	Area Absent	Area Present	Impulse Control	Chi Square	<i>p</i>
1. Marriage	8 2	10 16	High Low	5.0	.05
2. Parent-Child Relationship	13 4	5 14	High Low	9.0	.01
3. Social Pathology	16 6	2 12	High Low	11.7	.001
4. Physical Illness	14 9	4 9	High Low	3.0	<i>ns</i>
5. Mental Illness	16 9	2 9	High Low	6.4	.02
6. Financial Support	12 4	6 14	High Low	7.2	.01

TABLE 2

INCIDENCES OF MULTIPROBLEM FAMILY INDEX AREAS
FOR 18 HIGH AND 18 LOW IMPULSE
CONTROL SUBJECTS

Number of Areas Scored	Incidence of Multiproblem Family Index Areas						
	0	1	2	3	4	5	6
High Impulse Control (18)	6	4	3	2	2	1	0
Low Impulse Control (18)	0	1	1	2	6	7	1
Number of Children (36)	6	5	4	4	8	8	1

evidencing four or five dimensions of family pathology.

In Table 2, when boys of high and low impulse control are compared on the basis of low family pathology (0-3 areas) and high family pathology (4-6 areas), a chi square of 13.5 results. This division of the subjects on the total Multiproblem Family Index is significantly related to impulse control beyond the .001 level. Here, few family problem areas are associated with high impulse control, many family problem areas with low impulse control.

The relationship between the extent of family pathology, as measured by the Multiproblem Family Index, and the child's level of impulse control, is a rather striking one. These results become all the more revealing through a study of the individual cases. As inspection of Table 2 reveals, the six families evidencing none of the six areas of family pathology all had boys who were in the high impulse control group. On the other hand, the one family that scored on all the areas of family pathology had a boy in the low impulse control group. Significantly, this boy was ranked thirty-sixth on impulse control in a population of 36.

Analysis of those cases in Table 2 that did not adhere to the general direction is also quite revealing. One of the three boys with high impulse control, despite coming from a multiproblem family, was by far the oldest subject, 15 years 3 months. His high impulse control may reflect the significant relationship between impulse control and age ($\rho = .39$, significant at the .05 level). Of the four low impulse control boys who came from

families with few problem areas, one is an autistic schizophrenic, two have findings of organic brain pathology, and the last is an enigma not only in the present research but to the clinical staff of the Astor Home as well. Certainly in the cases of the boys with low impulse control concomitant with brain damage, one might expect low impulse control despite the lack of family pathology. This also might be true in the case of the autistic child.

DISCUSSION

Colvin's (1958) research on the effects of institutionalization upon the development of impulse control suggests that lack of opportunity for parental identifications may be a major factor in the inadequate development of impulse control. The present research suggests that contact, and probably identification, with pathological parental figures also hinders the development of impulse control.

The Multiproblem Family Index can be viewed as a measure of the parents' control of their *own* impulses. That is, in the context of family living, impulse control may well be reflected by the ability to sustain a marriage, refrain from impulsively striking a child, stay within the boundaries of acceptable social behavior, and manage family financial affairs so as not to need public assistance. Although this interpretation might be less appropriate for mental illness in the family, even here inadequate control of primitive instincts and impulses is frequently viewed as central to the problem of mental illness. However, physical illness in the family certainly cannot be interpreted as evidence of parental impulsivity. This is consistent with the finding that family illness was the only problem area not related to the impulse control level of the subjects.

This seemingly close association between the impulse control level of children and their parents is consistent with Redl and Wine-man's (1957) observations of *The Aggressive Child*. They note that the parents of these children present a picture of impulsivity no less striking than that of their children:

Seldom do we see children who have been so grossly and continuously exposed to traumatization on so many different levels. . . . With them benign experience is the exception, trauma the rule. From the things we know about them, it is this that emerges

as the most powerful theme in their prior experience (pp. 54-55).

SUMMARY

This study investigated the relationship of family pathology, as measured by Bloch and Behrens Multiproblem Family Index, and the level of impulse control in 36 children. The presence of family pathology was found to be significantly associated with low impulse control, the absence of family pathology with high impulse control. It was suggested that contact, and probably identification, with pathological parental figures hinders the development of impulse control.

REFERENCES

- BLOCH, D. A., BEHRENS, MARJORIE L., GUTTENBERG, H., KING, FRANCES G., & TENDLER, DIANA. *A study of children referred for residential treatment in New York State*. Albany: New York State Interdepartmental Health Resources Board, 1959.
- COLVIN, R. W. Defective ego and social development as functions of prolonged institutionalization of children. *Amer. Psychologist*, 1958, 13, 327. (Abstract)
- FREUD, S. The interpretations of dreams. In A. A. Brill (Ed.), *The basic writings of Sigmund Freud*. New York: Modern Library, 1938.
- GOLDFARB, W. Rorschach test differences between family reared, institution reared, and schizophrenic children. *Amer. J. Orthopsychiat.*, 1949, 19, 624-633.
- MELTZOFF, J., SINGER, J. L., & KORCHIN, S. J. Motor inhibition and Rorschach movement responses: A test of the sensory tonic theory. *J. Pers.*, 1953, 21, 400-410.
- RAPAPORT, D. *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951.
- REDL, F., & WINEMAN, D. *The aggressive child*. Glencoe, Ill.: Free Press, 1957.
- SCHILDER, P. On the development of thoughts. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951.
- SINGER, J. L., & HERMAN, J. Motor and phantasy correlates of Rorschach human movement responses. *J. consult. Psychol.*, 1954, 18, 325-331.
- SINGER, J. L., WILENSKY, H., & MCCRAVEN, VIVIAN. Delaying capacity, fantasy, and planning ability: A factorial study of some basic ego functions. *J. consult. Psychol.*, 1956, 20, 375-383.
- WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

(Received December 3, 1959)

EXAMINER REINFORCEMENT AND SITUATIONAL VARIABLES IN A PROJECTIVE TESTING SITUATION¹

LAWRENCE SIMKINS²

University of Houston

A clinical area which lends itself to the application of operant conditioning principles is the projective testing situation. Psychologists readily agree that there are wide individual differences in the behavior of examiners in a testing situation. A number of investigations (Alden & Benton, 1951; Baughman, 1951; Fahmy, 1953; Gibby, 1952; Lord, 1950; Sanders & Cleveland, 1953) have already shown that in many instances the responses given by a subject (*S*) are dependent upon his subjective definition of the test, personality characteristics and attitudes of the experimenter (*E*), and the conditions under which the test is administered. It is quite possible also that in the testing situation responses emitted inadvertently by the *E* are serving as reinforcers or punishment for some classes of the *Ss'* responses. Recent studies by Fahmy (1953), Wickes (1956), and Gross (1959) showed that it was possible to increase the frequency of human movement responses given by *Ss* in an inkblot testing situation through the application of reinforcement techniques.

This study was designed to answer the following questions: (a) What dimensions of a particular projective technique are most susceptible to examiner reinforcement, that is, content, location, or determinants? (b) What is the effect of obvious or strong approval as opposed to subtle or weak approval? Is weak approval any more effective than no approval

at all? (c) What is the effect of situational variables in a projective testing situation? Thus what effect will reinforcement have in situations that are perceived as ego involving as contrasted to more relatively neutral conditions?

METHOD

Study I

This study was designed to answer the first two questions. The third question required some modification in design and will be referred to as Study II.

The experimental situation involved the administration of an inkblot test devised by Holtzman (1958). The Holtzman cards were administered at each session. An immediate inquiry was conducted after each response. The *Ss* were college students selected from introductory psychology classes at the University of Houston and student nurses at the University of Texas Medical Branch. There was no particular selection procedure other than having approximately the same number of males and females due to the possibility of sex differences.

A group of *Ss* ($N = 25$) designated as the control group was administered 75 Holtzman cards over five testing sessions. Another group of *Ss* designated as the experimental group ($N = 36$) was administered 15 of the Holtzman cards in the first session. At the end of the first test session the experimental *Ss'* responses were scored and tabulated. There were three main dimensions considered in this research: content, location, and determinants. At this point the experimental *Ss'* protocols were analyzed to see if there were any other *Ss* who had given the same number of responses in this category. If there was such an *S*, both of these *Ss* were then matched and considered an experimental pair.

This pair of *Ss* was in turn matched with *Ss* from the general control group who had also given the same frequency of responses in this particular category in their first session. The *Ss* selected from the general control pool for matching purposes were referred to as the matched control group.

Procedure. Prior to the first testing session all experimental and control *Ss* were given the following instructions:

¹ This paper is derived from a study done by the author in partial fulfillment of the requirements for the PhD degree at the University of Houston. The author wishes to express his gratitude to the members of the doctoral committee for their help and encouragement in this work.

² Now at Florida State University.

You may have heard about the Rorschach inkblot test. It is a test used by clinical psychologists as a method for evaluating personality and the diagnosis of abnormal behavior. Unfortunately, many recent investigations conducted on the Rorschach have shown that it is not as effective as we would like it to be. Dr. Holtzman of the University of Texas has developed a new set of inkblots designed in such a manner that it is hoped will eliminate some of the weaknesses of the Rorschach test. Because of the comparative newness of this test we would like to get an idea of the types of things seen by the normal individual. *At this point we are not interested in personality factors* because we frankly do not yet know what the test measures. All we want to get is an idea of the common types of things that are seen. Also, unlike the Rorschach this test has many more inkblots so that it will be necessary to test you over a five-day period. Each testing should take only thirty to forty-five minutes. I am going to show you these cards one at a time. They are really only inkblots and people see different things in them. There are no right or wrong answers—you just tell me what you see. I would like you to see two things in every card—that is, give me two responses per card. Do you have any questions before we begin? O.K. Here's the first card.

Using this procedure the control Ss went through five testing sessions and no reinforcement was administered. At the end of the first session with the experimental Ss, the matching procedure described earlier took place. Beginning with the second session one experimental S of the matched experimental pair received a *strong* quality of reinforcement for each response he gave which was in the category that E had elected to reinforce. Strong quality reinforcement consisted of saying "good" or "that's fine." The other experimental S of this pair received a weak quality of reinforcement which consisted simply of saying "mm-hm" and nodding of E's head. Reinforcement was administered to the experimental Ss during the middle three sessions, that is Sessions 2, 3, and 4. On the fifth and final session reinforcement was not administered.

A total of 36 experimental Ss were employed in a three-by-two design with six replications. That is, there were three inkblots dimensions, two qualities of reinforcement, and six Ss in each cell.

Study II

The investigation designed to answer the third question of this research was performed as an independent study. However, in some respects it was dependent upon the results obtained in Study I. For one thing, whatever inkblot dimension (content, location, or determinants) proved to be the most susceptible to reinforcement was to be used in this study. Also, instead of using both strong and weak reinforcement, only weak reinforcement was used. This was a crucial aspect of this research in that it was the author's speculation that it is weak reinforcement

which is probably used routinely and inadvertently in the clinical testing situation. This is particularly relevant when one considers the factor of the anxiety of the patient in the testing situation. It is possible that under anxiety producing conditions the patient may become more sensitive to relatively subtle stimuli than he would under more neutral conditions.

The control group for this study received instructions that were made as neutral as possible with the emphasis being placed on a "new test we didn't know too much about" and minimizing personality factors.

The experimental Ss in this part of the study had not been used previously. There were two experimental conditions in this part of the experiment; Ss in both groups received weak reinforcement. One experimental group was run under conditions of a positive set; the other under a negative set. Although both sets were designed to induce ego involvement in the Ss, it was hoped that the positive set would motivate an S to "do his best" whereas it was expected that a negative set would provide a threatening test atmosphere. Again there was a three-way match on some *content* category between the two experimental Ss and a control S (or group of Ss). All of these Ss were matched on the basis of their first testing session. Unlike Study I, both experimental Ss received weak reinforcement. One experimental S of the pair received reinforcement under a negative set instruction; the other S under a positive set instruction.

A total of 12 experimental Ss (that is, six pairs of experimental Ss) participated under the two experimental set instructions. As in Study I, the Ss were tested over five sessions with no reinforcement being administered in the first and final sessions.

Prior to the first session, the experimental Ss received instruction similar to those given to the Study I Ss. However, the instructions were slightly modified to the extent of emphasizing that the first test session was only a practice session. Thus they were told:

Today's session will constitute a *practice session* which will serve to acquaint you with the inkblots. In other words I'm not particularly concerned with what you see today.

Following the first session (control) the S's protocol was evaluated and he was matched with a control and another experimental S on the particular class of responses the E wished to reinforce. He was then either assigned to a positive set or a negative set experimental group.

Negative Set. On the day of the second session, experimental Ss in the negative set group were given test instructions worded in a negative tone. Thus they were told:

These inkblots have been found adequate for indicating neurotic people The immediate purpose of this research is to determine the *neurotic potential* that is inherent within a *normal population*.

Under this experimental set, it was anticipated that such signs of interest on the part of the E would

TABLE 1

COMPARISON OF MEAN *D* SCORES OF EXPERIMENTAL SUBJECTS TO MEAN *D* SCORES OF THEIR RESPECTIVE MATCHED CONTROLS
(*N* = 6 in each group)

	Mean	<i>T</i> ^a
Location Reinforcement	1.8	3
Matched Control	-1.3	
Determinant Reinforcement	4.3	0*
Matched Control	-2.6	
Content Reinforcement	7.9	0*
Matched Control	-0.9	

^a Wilcoxon sign rank test.
* *p* = .025, one-tailed test.

constitute punishment. Again, as in previous investigations, reinforcements were not given in the final session.

Positive Set Instructions. The instructions given to this group on the first day were identical to those given to the experimental *Ss* in the negative set on their first day.

Positive Set. On the day of the second testing session the instructions to this group were modified in such a manner as to create a "positive" test atmosphere. Thus, the *Ss* were told:

The inkblots you will be shown from now on have been given to a superior intellectual group composed of people that have been successful in their professions Despite the divergent interests (of these people) the tests get at what we refer to as a common *creative potential* The imme-

TABLE 2

COMPARISON OF *D'* SCORES OF WEAK AND STRONG QUALITY REINFORCEMENT
(*N* = 18 in each group)

	Mean	<i>SD</i>	<i>T</i> ^a
Weak	6.3	4.8	
Strong	5.8	5.9	0.33*

^a Wilcoxon pair rank test.
* *p* = .40.

diate purpose of this research is to determine just how much of this intellectual and creative potential you possess.

The reinforcement for *Ss* in this group was identical to the experimental *Ss* in the negative set group. It was hoped that under this experimental condition the responses emitted by *E* would serve as reinforcing stimuli to increase the frequency of a certain class of responses.

RESULTS

Study I

The change or the difference in the number of responses in a reinforced category from the first to the last session for a given *S* from now on will be referred to as a *D* score. A minus sign before this score indicates a decrease, the omission of a sign indicates an increase.

A comparison was made between the *D* scores of the experimental *Ss* and the *D* scores

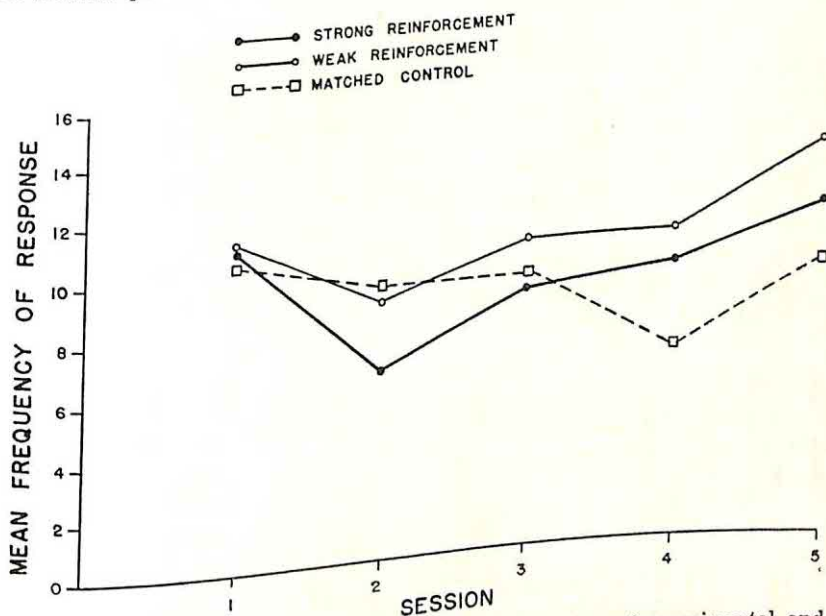


FIG. 1. Mean frequency of responses in each test session of experimental and matched control subjects in the location dimension.

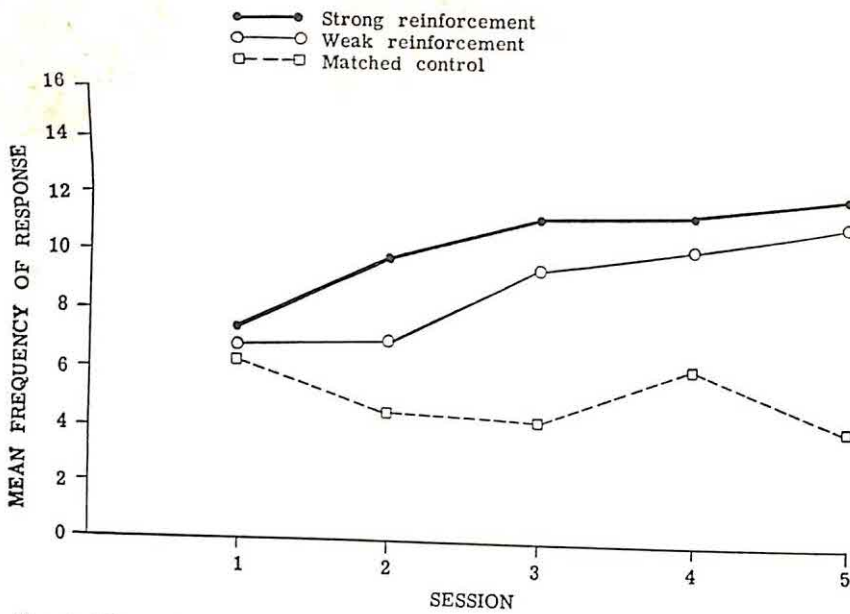


FIG. 2. Mean frequency of responses in each test session of experimental and matched control subjects in the determinant dimension.

of their respective matched controls for each dimension. Using the Wilcoxon matched pair sign rank test, the results in Table 1 list the mean D scores for experimental and matched control S s within the respective inkblot dimensions. Only the location dimension failed to reach significance.

In order to compare the effects of reinforcement

quality a different score was derived. The D' score represents the net change or the difference between D score of the experimental S and his respective matched control. Using the D' scores, each strong quality reinforcement S was compared with the matched experimental S that received weak quality reinforcement. The results presented in Table 2

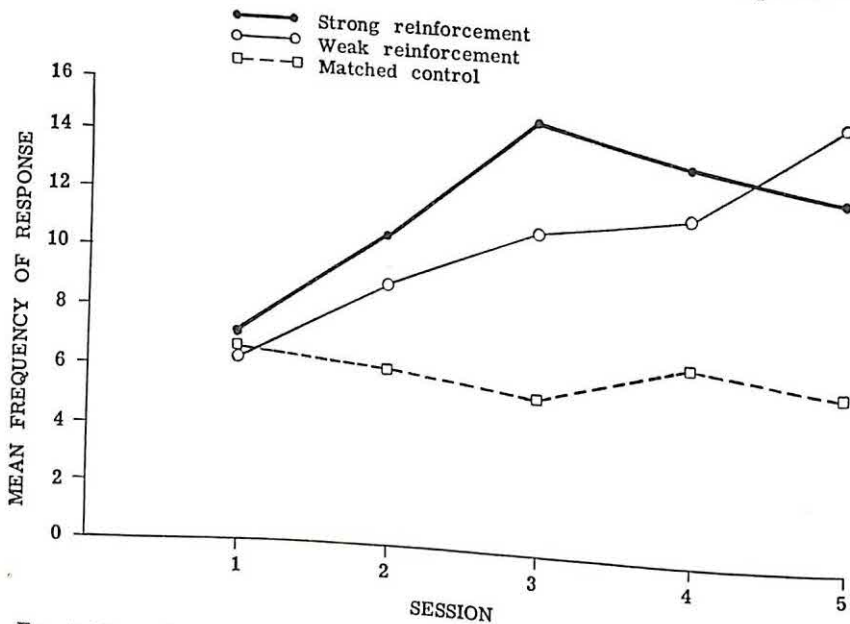


FIG. 3. Mean frequency of responses in each test session of experimental and matched control subjects in the content dimension.

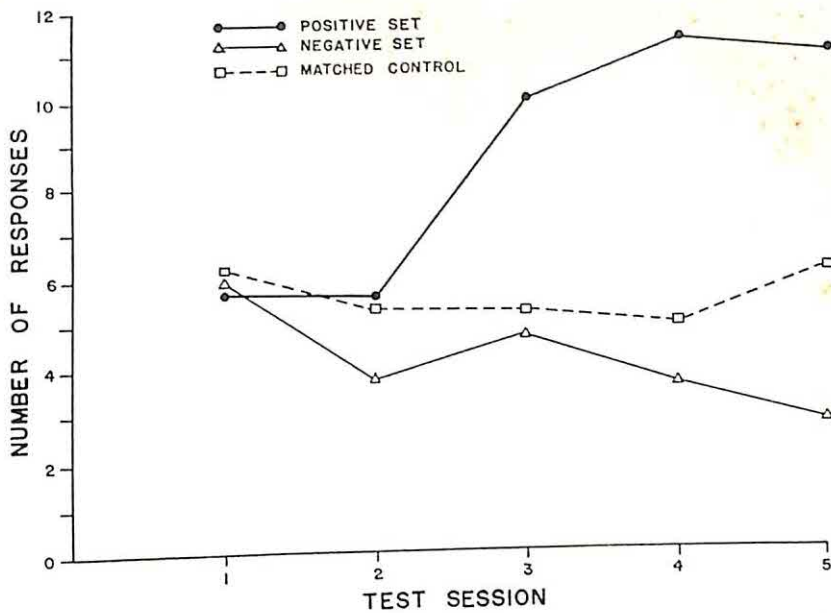


FIG. 4. Mean number of responses independent of the reinforcement category given at each test session by experimental subjects and their matched controls in Study II.

show that the differences between qualities of reinforcement were not significant for 17 degrees of freedom.

The graphs presented in Figures 1, 2, and 3 show the mean number of responses given in the reinforced categories by experimental Ss and their matched controls for each testing session. In both the location and content dimensions weak reinforcement was slightly more effective than strong reinforcement. The reverse was true of the determinant dimension. In the determinant dimension, however, one experimental S receiving strong reinforcement showed an unusually large increase in the reinforced category, thus producing an artificial elevation of the mean for that group. If his score is deleted and the mean of only five Ss used, then the weak reinforcement

mean would be slightly higher than the strong reinforcement. However, there were no significant differences between weak and strong quality reinforcement in any of the inkblot dimensions.

Study II

Again using D' scores, a comparison of the two experimental groups (positive set and negative set) in Study II was made. The mean D' scores for each experimental group are shown in Table 3. The graph presented in Figure 4 shows the mean number of responses given in each session by the experimental Ss and their matched control Ss. Since one of the problems of this research was to study the relative effectiveness of reinforcement in an anxiety producing situation, a comparison was made between the D' scores of content Ss receiving weak reinforcement in Study I and the D' scores of the positive set Ss of Study II. The mean increase of the Study I Ss was 9.5; the mean increase of the Study II Ss was 7.3. The differences were not significant.

DISCUSSION

Study I

The results show that the content and determinant dimensions are about equally vulnerable to reinforcement effects. It is also

TABLE 3
COMPARISON OF MEAN D' CHANGES OF POSITIVE AND NEGATIVE SET EXPERIMENTAL SUBJECTS
($N = 6$ in each group)

	Mean	T^a
Positive Set	+7.3	0*
Negative Set	-3.0	

^a Wilcoxon pair rank test.
* $p = .025$, one-tailed test.

fairly clear that of the three inkblot dimensions, the location dimension is the most difficult to manipulate, that is, shows the least change as a function of reinforcement. One possible reason for the ineffectiveness of reinforcement on the location category may lie in the nature of the response class. The location dimension is vaguely defined, in that there is a great deal of dissimilarity between the various members that compose this response class. Because of the vagueness of this category and the relatively short training period with the experimental Ss, an adequate stimulus discrimination is difficult to achieve. Previous investigators (Buss & Durkee, 1958; Greenspoon, 1955; Salzinger, in press) have considered the nature of the response class an important variable in determining the effects of reinforcement. In general, the larger the response class, the less effective is the conditioning. As the size of the response class increases, there is frequently a corresponding increment in heterogeneity among members of the response class. Consequently, generalization of reinforcement to all members of the response class is not as effective as with a more homogeneous class.

Although there is no difference between weak and strong quality reinforcement it was somewhat surprising to find that weak reinforcement was just as effective and in some cases more effective than strong reinforcement. It is interesting to note that similar results were found by Wickes (1956) and Gross (1959). Both of these studies were concerned with investigating the effects of verbal versus nonverbal reinforcement in conditioning responses on an inkblot test. In both studies verbal and nonverbal reinforcements were effective with nonverbal being slightly more effective than the verbal reinforcement.

The results of this study indicate that expressions of "neutral" interest that the clinician uses in administering a test can serve as reinforcing stimuli.

Study II

The purpose of Study II was to determine the effects of reinforcement under experimental conditions that were designed to induce tension in the Ss. It was expected that the same "weak" stimuli that *E* made contingent

upon a given class of responses would have different effects under these two experimental sets. As can be seen from Figure 4 and Table 3, the two experimental groups responded to the reinforcing stimuli in a different manner. Although the amount of change (*D*) was not significant in every case, the direction of change was consistent with all experimental Ss.

The small mean amount of change (-3.0) on the part of the negative set group was probably a function of a limitation in the experimental design. Classes of responses containing five to eight responses in the initial session were considered as probably the best for experimental manipulation. This range of responses was found to be satisfactory for demonstrating the effects of positive reinforcement; it is not adequate, however, for effectively demonstrating a decrease. Even if the *S* drops from five responses in his initial session to no responses, the decrease is not very striking. In light of these considerations, any conclusions regarding the effectiveness of reinforcement under conditions of a negative set are, at best, quite tenuous. Future research employing a better design and a larger number of Ss is necessary to substantiate these findings.

The interpretation of Study II results becomes even more difficult when a comparison is made between the neutral set Ss of Study I and the positive set Ss of Study II. The difference between these two groups was not significant. Does the induction of anxiety have very little to do with reinforcement effect? A partial answer is suggested by some of the verbalizations of the experimental Ss in Study II who stated that although they felt very uncomfortable and nervous during the earlier stages of the testing they felt less anxious as the testing progressed. Apparently the *E*'s attempt to induce anxiety was somewhat unsuccessful. One explanation of this failure is that there may be adaptation effects in a prolonged testing series. It will be recalled that the anxiety producing instructions were given in the second session and not mentioned again after that. As the sessions progressed the Ss might have experienced less subjective discomfort. The *E* also noticed that in the earlier stages of the testing the Ss seemed to make more affective statements concerning the ink-

blots, and expressed more concern about their performance than in later sessions.

In view of the questionable effectiveness of the anxiety producing set, definitive conclusions cannot be drawn. Future research along these lines should make an attempt to maintain a high anxiety level if a prolonged testing series is required. One trend from Study II does appear to be evident. There is an interaction between what the *S* conceived the test to be and the manner in which stimuli serving as reinforcement affected some of his responses.

SUMMARY

The purpose of this research was to study the behavioral effects of certain contingent stimuli emitted by the examiner in a projective testing situation. This research was divided into two separate investigations. The first investigation was concerned with determining what dimensions of an inkblot test (content, location, determinants) were the most sensitive to reinforcement and also comparing the effects of strong and weak quality reinforcement. Strong quality reinforcement consisted of the examiner's saying "good" or "that's fine" to a particular response in the class of responses the examiner had decided to reinforce. Weak quality reinforcement consisted of the examiner's nodding his head and saying "mm-hm." This part of the research was referred to as Study I.

The second part of the investigation was concerned with situational variables and their relationship to examiner reinforcement. Thus it was expected that under anxiety producing situations the effects of examiner reinforcement would be accentuated. This part of the research was referred to as Study II.

The results of Study I were as follows:

1. The content and determinant dimensions showed the largest reinforcement effects.
2. The location dimension was the most resistant to change as a function of reinforcement.
3. There was no difference between strong and weak reinforcement; if anything, weak reinforcement was slightly more effective than strong reinforcement.

The results of Study II were as follows: There were significant differences between the negative and positive set experimental subjects. Under positive set instructions weak quality stimuli had the effects of positive reinforcers. Under negative set instructions weak quality stimuli had a punishing effect producing a decrease in the reinforced category. Only tentative limitation conclusions with respect to the effectiveness of reinforcement under a negative set, however, could be drawn because of design and the few number of subjects employed. This was attributed to the questionable effectiveness of the anxiety producing instructions given only at the beginning of a long testing series.

REFERENCES

- ALDEN, P., & BENTON, A. Relationship of sex of examiner to incidence of Rorschach responses with sexual content. *J. proj. Tech.*, 1951, 15, 231-234.
- BAUGHMAN, E. Rorschach scores as a function of examiner differences. *J. proj. Tech.*, 1951, 15, 243-249.
- BUSS, A., & DURKEE, A. Conditioning of hostile verbalizations in a situation resembling a clinical interview. *J. consult. Psychol.*, 1958, 22, 415-418.
- FAHMY, S. Conditioning and extinction of a referential verbal response class in a situation resembling a clinical diagnostic interview. *Dissertation Abstr.*, 1953, 13, 873-874.
- GIBBY, R. Examiner influence on the Rorschach inquiry. *J. consult. Psychol.*, 1952, 16, 449-455.
- GREENSPOON, J. The reinforcing effect of two spoken sounds in the frequency of two responses. *Amer. J. Psychol.*, 1955, 68, 409-416.
- GROSS, L. Effects of verbal and nonverbal reinforcement on the Rorschach. *J. consult. Psychol.*, 1959, 23, 66-68.
- HOLTZMAN, W. *Holtzman Inkblot Test*. Austin: Univ. Texas Press, 1958.
- LORD, E. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, 64(10, Whole No. 316).
- SALZINGER, K. Experimental manipulation of verbal behavior. *J. gen. Psychol.*, in press.
- SANDERS, R., & CLEVELAND, S. E. The relationship between certain examiner personality variables and subjects' Rorschach scores. *J. proj. Tech.*, 1953, 17, 34-50.
- SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.
- WICKES, T. Examiner influence in a testing situation. *J. consult. Psychol.*, 1956, 20, 23-26.

(Received December 11, 1959)

FURTHER EVIDENCE CONCERNING THE EFFECT OF HANDWRITING HABITS UPON THE LOCATION OF DRAWINGS¹

WAYNE DENNIS AND EVELYN RASKIN

Brooklyn College

The placement on the page of a human figure drawing often is interpreted as resulting from personality factors (Levy, 1950; Machover, 1949). Although the location of drawings *may* be related to personality, other influences are also operative. Among these is the influence of handwriting habits upon drawings. In a previous study (Dennis, 1958) it was found that among American, Armenian, and Arab subjects (Ss), drawings are located most frequently in that quadrant of the page within which the S ordinarily begins to write. This is the top-left quadrant for Americans and Armenians and the top-right quadrant for Arabic-writing peoples.

This finding is not entirely conclusive since the differences between the groups in regard to the location of drawings might conceivably be due to factors other than handwriting. However, if in many other linguistic groups, the location of drawings is found to be related to the starting position in handwriting, the hypothesis that handwriting habits affect drawing responses will be strengthened. The present paper presents data concerning the location of drawings in the case of five additional linguistic groups, three of which begin writing at the top-left of the page and two of which begin at the top-right of the page.

METHOD

The top-left handwriting groups consisted of Turks, Cambodians, and Japanese. The top-right handwrit-

¹ The collection of the data on which the present paper is based was assisted by a grant to the senior author by the Social Science Research Council. For making possible the collection of the data we are indebted to the following persons: in Turkey, Omer Saray; in Iran, Abbas Hadian and Zahra Dideban; in Japan, Koji Sato and Minoru Hirota; in Cambodia, Carleton Washburn, Helen Brell, and Ho Tong Ho; in Israel, May Bere Merom.

ing groups consisted of Iranians and Israelis. The number of Ss in each group is indicated in Table 1.

In each group the Ss were school children aged 6-12. They were given either the Goodenough Draw-a-Man Test or the Machover Draw-a-Person Test. Wherever possible, drawings were obtained in pencil on 8½" by 11" white unruled paper.

As in the previous study, each drawing was scored as to whether the topmost or bottommost part of the drawing was nearer to its respective edge of the drawing paper. Similarly, we determined which side of the body was nearer to its corresponding lateral edge of the paper. This produced a four-fold categorization of locations: top-left, top-right, bottom-left, and bottom-right. While this method of scoring does not attempt to measure the body surface lying in each quadrant, in the majority of instances, the scored quadrant contains a larger part of the area of the drawing than does any other quadrant.

RESULTS

Table 1 shows that the results of this study are completely in accord with the motor transfer hypothesis. In each of the five linguistic groups, the quadrant which contains the largest number of location scores is that in which the initial handwriting movement is

TABLE 1
LOCATION OF DRAWINGS IN DIFFERENT
HANDWRITING GROUPS

DRAWING GROUPS					
Group	N	Location of Drawings (%)			
		T-L	T-R	B-L	B-R
Top-Left					
Turkish	289	42	21	22	14
Cambodian	138	44	9	25	22
Japanese	349	38	16	34	12
Top-Right					
Iranian	307	16	40	17	27
Israeli	366	27	45	12	16

ordinarily made. Computation shows that in each group the inequality of the quadrants exceeds the .001 level of confidence.

The above finding should not be construed as necessarily indicating that placement of the drawing on the page is devoid of significance for personality interpretation when the writing habits of the individual child are known. Within each of our groups, as Table 1 shows, there were always some children who placed their drawings in positions which deviated markedly from their writing conventions. Such deviations may indicate differences in a child's personality. Our data do not show whether this is, or is not, the case.

SUMMARY AND INTERPRETATION

Corroborating the results of an earlier study which involved three linguistic groups, the present study of five additional linguistic

groups indicates that human figure drawings tend to be located in that quadrant of the page within which the subject begins to write, to a greater degree than in any other quadrant.

Thus this study, like the preceding one, supports the hypothesis that a factor other than personality plays a marked role in creating intergroup differences in locations of drawings.

REFERENCES

- DENNIS, W. Handwriting conventions as determinants of human figure drawings. *J. consult. Psychol.*, 1958, 22, 293-295.
- LEVY, S. Figure drawings as a projective test. In L. E. Abt & L. Bellack (Eds.), *Projective psychology*. New York: Knopf, 1950. Pp. 257-297.
- MACHOVER, K. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.

(Received October 2, 1959)

ACHIEVEMENT MOTIVATION AND FIELD INDEPENDENCE

GILBERT HONIGFELD AND IRWIN M. SPIGEL¹

Temple University

Wertheim and Mednick (1958) have pointed out the similarities between McClelland's analysis of the origins of achievement motivation (McClelland, 1951; McClelland, Atkinson, Clark, & Lowell, 1953) and Witkin's treatment of the developmental antecedents of field independent modes of perception (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954). These two positions attribute low need for achievement (*n* Ach) and perceptual field dependence to restrictive parental pressures and few demands for independent behavior in childhood. For this reason Wertheim and Mednick (1958) predicted a significant positive relationship between *n* Ach and field independence when measured in adulthood. Using 31 female and 11 male undergraduates they obtained a .40 product-moment correlation.

Owing to the possible differences between males and females in terms of their performance on these variables a more balanced sample was deemed desirable, and so this replication was performed. The basic hypothesis of a significant positive relationship between *n* Ach and field independence remained unaltered, but additional efforts were undertaken to explore the possibility of systematic sex differences.

Fifteen male and 13 female subjects (*Ss*) were used. They were undergraduates at a large, urban university and were naive regarding the instruments used and the goals of the study. *Ss* in small numbers were administered under neutral set a group adaptation of the McClelland et al. (1953) four card TAT for measuring *n* Ach. They were indi-

vidually administered a modification of the Gottschaldt Embedded Figures Test (EFT), in which simple figures had to be isolated from complex figures (Thurstone, 1944).

Scores on the *n* Ach test ranged from -2 to +19 with a median of 8.5 (the higher the score, the greater the achievement motivation). All scoring was done by one of the authors who obtained a rescoring reliability coefficient of .92. Total times to isolate simple from complex figures ranged from 216" to 1660" with a median of 703.5". Both males and females were shown to be homogeneous with regard to median scores on both variables (chi square tests indicated no differences).

A Yates-corrected chi square of 6.04 ($p < .05$, 2 *df*) was obtained between EFT scores dichotomized at the median and the *n* Ach test trichotomized at the upper and lower quartiles, after the manner of McClelland et al. (1953). Peculiarities in the distributions were suggested when a 2×2 chi square analysis (median test) on the same data proved statistically insignificant. It appeared that such peculiarities might be attributable to sex differences. When separate rank-order correlations were computed for each sex it was found that only the females behaved in the manner suggested by the Wertheim and Mednick (1958) hypothesis. The degree of correlation for the women (.42, 11 *df*) agrees with their earlier finding. In the case of the males a slight correlation (-.12, 13 *df*) in the opposite direction was revealed. The difference between correlations was significant beyond the .01 level. In light of these apparently divergent trends, the low-order overall correlation obtained (.24, 26 *df*) suggests

¹ The help of Horace A. Page in supplying materials and assistance is appreciated.

an interaction effect operating across sexes. It will be recalled that the Wertheim and Mednick (1958) sample favored females in the ratio of 3 to 1. It is possible then that their results reflect the female contribution which alone is in the hypothesized direction.

In conclusion, the replication confirmed the finding by Wertheim and Mednick (1958) of a significant positive relationship between *n* Ach and field independence. However, there is evidence in the present study to suggest that this hypothesis is pertinent only to a female population.

REFERENCES

- McCLELLAND, D. C. *Personality*. New York: Sloane, 1951.
- McCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- THURSTONE, L. L. *A factorial study of perception*. Chicago: Univer. Chicago Press, 1944.
- WERTHEIM, J., & MEDNICK, S. A. The achievement motive and field independence. *J. consult. Psychol.*, 1958, 22, 38.
- WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

(Received October 23, 1959)

RATIONALES FOR THE "TWISTED PEAR"

LOWELL H. STORMS

University of California Medical Center, Los Angeles

In a recent article, Jerome Fisher (1959) described a phenomenon which he termed the "twisted pear." Essentially, this refers to the fact that in a number of situations in which a predictor variable with a cutting point is related to a dichotomous criterion such as brain damage present or not present—or poor versus good adjustment—the predictions made yield a substantial proportion of true positives at the expense of very few false positives, but with a large number of false negatives. This phenomenon appears to be more than an artifact of the position of the cutting score of the predictor, since it is not possible to discriminate false negatives from true negatives as effectively by varying the cutting score.

It is the purpose of the following paragraphs to show: (a) that the occurrence of a twisted pear does not imply that normal distributions are not characteristic of the populations involved, (b) that the twisted pear is not universal if defined in such a way that it is not an artifact of placement of the cutting score on the predictor, (c) that various relationships between predictors and criteria can result in twisted pears.

By quoting a statement made by Boring, Fisher at least implies that his twisted pear is inconsistent with a normal curve. However, the figures in his fourfold tables are derived from considering two distributions, each divided in two parts by a cutting score on the predictor. Especially when the abnormal population is heavily overrepresented with respect to the base rate of the abnormality in the general population, two normal distributions could easily combine to produce a twisted pear. This is exactly what would happen if the mean of the abnormals on the predictor were higher than that of the normals and the variance were also greater in such a way that the range of the abnormals' distribution began

at the bottom of the normals' distribution but extended further into the "high" scores.

The relationship between IQ as a predictor variable and educational achievement as a criterion is a poor example of the twisted pear. While low IQs accurately identify persons who will not go to college (with few false positives and many false negatives) high IQs in Fisher's table do the same thing for people who will achieve at least at the ninth grade level. Tables 1 and 2, which are derived from the table on page 402 in Fisher's article (1959), show that equally effective predictions can be made using both high and low cutting scores.

Since discrimination can be obtained at either end of the continuum merely by shifting the cutting point on the predictor, Fisher's description of the twisted pear does not apply. This clearly indicates that the generality of the phenomenon he describes is not unlimited.

One set of circumstances which can yield twisted pear predictor-criterion fourfold tables has to do with the multiplicity of causes bringing about the poor performance or inadequate adjustment represented by one part of the dichotomous criterion. These causes may function in an interrelated manner or may be relatively independent. If we assume perfect reliability of our criterion and perfect reliability and validity of a predictor as a measurement of just one or a subset of several of the

TABLE 1
RELATIONSHIP BETWEEN LOW INTELLIGENCE
AND EDUCATIONAL ACHIEVEMENT

Intelligence	Achievement	
	No College	Some College
Mental defective	100%	0%
Control	71%	29%

TABLE 2
RELATIONSHIP BETWEEN HIGH INTELLIGENCE
AND ACHIEVEMENT

Intelligence	Achievement	
	Grade 8 and lower	Grade 9 or more
Gifted	0%	100%
Control	36%	64%

causes related to a malfunctioning of the organism, it is to be expected that a subgroup of those individuals who are nonadaptive with respect to the criterion will be *accurately* (no false positives) picked up by the predictor. Other predictors related to other causes would be needed to identify cases missed by the predictor in question. Each predictor can be considered as a way of dipping into a pool (class) of unfortunates in order to select a subclass of problem cases. In this way, each simple or complex predictor would produce a twisted pear effect with respect to the criterion because some cases would be accurately identified and others left behind.

Consider for a moment people with hearing difficulties and assume we cannot ask them about their hearing. Relatively independent causes of hearing loss can consist of defects in the eardrum, middle ear, cochlea, and central nervous system. A good determination of the integrity and functioning of the eardrum would pick up some hearing problems without false positives but with many false negatives.

In a similar fashion, other hearing problems would be discovered by investigating the bones of the middle ear or the functioning of the cochlea, but in each case many hearing problems would be missed.

Clearly, predictor variables need not be determinations of causes but often consist of measurements of concomitant effects. For example, psychological consequences of brain damage such as memory disturbances, Piotrowski's signs, or impairment in perceptual-motor coordination, or physiological and anatomical anomalies such as EEG disturbances or displacement of the ventricles will identify different cases of damage to the central nervous system. Even though there will be overlap in the sense that some of the cases will be identified by more than one measure, the fact that some cases are missed by each measure yet identified by other means will make the predictor-criterion relationship look like a twisted pear for that measure.

More complex circumstances and relationships than those represented by the above examples could also produce twisted pears. The phenomenon is neither universal nor unitary and depends upon a variety of conditions affecting the distributions of both predictor and criterion measurements.

REFERENCE

- FISHER, J. The twisted pear and the prediction of behavior. *J. consult. Psychol.*, 1959, 23, 400-405.

(Received December 15, 1959)

BRIEF REPORTS

MANIFEST ANXIETY AND SOCIAL PERCEPTION

JOEL R. DAVITZ

Teachers College, Columbia University

AND

DONALD J. MASON

University of Colorado

This study investigated the relationship between manifest anxiety and two aspects of social perception: accuracy of perception and assumed similarity. Previous studies report somewhat contradictory findings concerning the relationship of adjustment and accuracy of perception, although the results of both studies support the hypothesis that personal adjustment is positively related to the tendency to perceive others as similar to oneself.

The subjects (Ss) of this research were 156 members of three college fraternities. Each S answered an interest questionnaire for himself and predicted the answers of six other fraternity members whose names followed the S's name on an alphabetical list of his fraternity. The interest questionnaire listed 35 activities, such as playing basketball, solving mathematical puzzles, and singing in a glee club. The S indicated his interest in each activity on a five-point scale ranging from *very interested* to *definitely do not like*.

Each S also completed the short form of the Manifest Anxiety Scale developed by Bendig (1956). The data were analyzed by comparing Ss in the upper (high anxious) and lower (low anxious) quartiles of the distribution of manifest anxiety scores.

Following the general procedure used by Bieri, Blacharsky, and Reid (1955) and also by Chance (1958), accuracy of perception scores was computed separately for items on which Ss were actually similar to the persons whose responses they predicted, and for items on which they were actually dissimilar on the interest questionnaire. Answers to a given item were considered *similar* if they did not differ by more than one scale-unit on the five-point scale indicating interest in each activity; answers were defined as *dissimilar* if they differed by more than one scale-unit. Ac-

curacy scores were obtained by summing the arithmetic differences in ratings of an S's predictions of the six persons whose responses he predicted, and the actual responses of these six persons on the interest questionnaire. Assumed similarity scores were computed by summing the arithmetic differences in ratings of an S's own responses and his predictions of the responses of the six persons whose answers he predicted.

There was no significant difference between the high and low anxious groups for either accuracy of perception on similar items, or accuracy of perception on dissimilar items. However, the difference between the high and low anxious groups for assumed similarity was significant at the .01 level in the direction of high anxious Ss perceiving themselves as less similar to others.¹

The results for assumed similarity are consistent with the two previous studies (Bieri et al., 1955; Chance, 1958); however, the absence of significant differences for accuracy of perception appears to contradict the findings of Bieri, Blacharsky, and Reid.

REFERENCES

- BENDIG, A. W. The development of a short form of the Manifest Anxiety Scale. *J. consult. Psychol.*, 1956, 20, 384.
BIERI, J., BLACHARSKY, E., & REID, J. W. Predictive behavior and personal adjustment. *J. consult. Psychol.*, 1955, 19, 351-356.
CHANCE, JUNE E. Adjustment and prediction of others' behavior. *J. consult. Psychol.*, 1958, 22, 191-194.

(Received November 5, 1959)

¹ A summary of the statistical analysis of the data and a copy of the questionnaire may be obtained from the senior author.

THE INTERPRETATION OF THREE PERSONALITY FACTORS IN JUVENILE DELINQUENCY¹

HERBERT C. QUAY,
Vanderbilt University

DONALD R. PETERSON,
University of Illinois

AND CONRAD CONSALVI
Vanderbilt University

In a previous study (Peterson, Quay, & Cameron, 1959) a factor analysis of two personality scales known to measure tendencies toward juvenile delinquency revealed the presence of three personality and two environmental background factors. The three personality factors were tentatively labeled "Psychopathic Delinquency" (PD), "Neurotic Delinquency" (ND), and "Inadequate Delinquency" (ID).

This study sought further clarification of the meaning and validity of these three personality factors by investigating their relationships to variables logically related to them.

Two groups of subjects (Ss) were used. The first group consisted of 170 consecutive admissions to the Illinois State Training School for Boys while the second group was composed of 93 boys incarcerated in the Tennessee State Vocational Training School. Approximately one-half of this latter group were new admissions whereas the remainder had been institutionalized for periods of one month to one year. Both groups were administered the factor scales, while the Tennessee sample was administered, in addition, the children's form of the Taylor Manifest Anxiety Scale (CMAS).

For Group 1, scores on each of the three factor scales were related, by appropriate correla-

tional methods, to: (a) type of offense leading to commitment, (b) problem behavior in the institution, (c) length of incarceration, (d) recidivism, and (e) intelligence as measured by the Otis test. For the second group, in which information in regard to the above variables was unavailable, factor scores were related only to CMAS.

Results indicated that the PD factor was significantly related to commitment for crime against the person ($r_{pbi} = .14$, $p = .05$), recidivism ($r_{pbi} = .16$, $p = .05$), problem behavior while incarcerated ($r_{pbi} = .16$, $p = .05$), and longer institutionalization ($r = .19$, $p = .01$), but was not significantly related to intelligence or CMAS. The ND factor related significantly only to CMAS ($r = .52$, $p = .01$). ID related significantly to intelligence ($r = -.23$, $p = .01$) and to longer institutionalization ($r = .16$, $p = .05$).

While the correlations were rather uniformly of low magnitude, they were, with but one exception, in keeping with the generally accepted meanings of the factors. The unexpected relationship of ID to length of institutionalization suggests a failure to adjust to institutional routine based on ineptness rather than a pattern of rule-violating psychopathy.

The results are interpreted as providing empirical support for the applicability of factor labels originally arrived at through analysis of item content.

REFERENCE

- PETERSON, D. R., QUAY, H. C., & CAMERON, G. R. Personality and background factors in juvenile delinquency as inferred from questionnaire responses. *J. consult. Psychol.*, 1959, 23, 395-399.

(Received November 9, 1959)

¹ An extended report of this study may be obtained without charge from H. C. Quay (Department of Psychology, Vanderbilt University; Nashville, Tennessee) or for a fee from the American Documentation Institute. Order Document No. 6407 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

MANIFEST ANXIETY AND SOCIAL BEHAVIOR¹

JOEL R. DAVITZ

Teachers College, Columbia University

Although various theories of anxiety have emphasized the relationship between anxiety and negatively evaluated social behavior, relatively little research has been focused on this problem. The purpose of this study, therefore, was to test the hypotheses that persons who report high anxiety, in comparison to those who report low anxiety, are described by others as more withdrawn, inhibited, hostile, unconscientious, conceited, and selfish.

The procedure involved collecting the following data: (a) self-reports of anxiety, (b) ratings of the social behavior of each subject (*S*) by persons with whom each *S* lived. The hypotheses were tested by comparing the ratings by others of *Ss* in the upper and lower quartiles of the total sample on the scale of manifest anxiety.

The *Ss* were 157 male college students who were members of three fraternities at a large midwestern university. Initial testing consisted of group administration of a short form of the Manifest Anxiety Scale (Bendig, 1956). Four to six weeks later, social behavior ratings were obtained. Each fraternity was divided randomly into four subgroups, with from 11 to 15 fraternity brothers in each subgroup. Each *S* rated

the fraternity brothers assigned to his particular subgroup on six eight-step scales: friendly-withdrawn, frank-inhibited, good-natured-hostile, conscientious-unconscientious, modest-conceited, generous-tight. Brief definitions were given of each point on the scales.

The distribution of social behavior ratings was skewed toward the positively evaluated end of each scale; however, the high anxious *Ss*, in comparison to the low anxious *Ss*, were described by their peers as less frank, less good-natured, less conscientious, and less generous. The differences on these scales were significant beyond the .05 level. The differences between the high and low anxious groups were not significant on two scales: friendly-withdrawn, and modest-conceited.

In general, high anxious *Ss* were described less positively than were low anxious *Ss*. Theoretical interpretation of these results, however, are limited by the fact that the data essentially represent correlations between reported anxiety and ratings of social behavior. Obviously, an antecedent-consequent relationship between anxiety and social behavior cannot be inferred; nevertheless, this research supports the general thesis that manifest anxiety and various kinds of social behavior covary, and the task of future research is to clarify the nature of these relationships.

REFERENCE

- BENDIG, A. W. The development of a short form of the Manifest Anxiety Scale. *J. consult. Psychol.*, 1956, 20, 384.

(Received November 19, 1959)

¹ An extended report of this study has been deposited with the American Documentation Institute. Order Document No. 6408 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

AN EMPIRICAL EVALUATION OF A TEST OF ROLE-PLAYING ABILITY¹

ZANWIL SPERBER

Children's Hospital, Philadelphia

Despite the importance of the role-playing ability (RPA) concept and related concepts (empathy, social sensitivity, etc.) little empirical research has been done relating RPA to other personality variables; to performance in relevant professions (e.g., psychotherapist); or to differences in communication, perception, and other functional skills. Hitherto, most researchers on these variables have used a clumsy cross-prediction design, where the need for comparison data restricts applications to a relatively small number of Ss who have had an opportunity to observe each other. McClelland (1951) recognized the potential value of a more generally applicable measure and developed a 32-item questionnaire as a test of RPA. It is the purpose of the present study to replicate McClelland's developmental procedure and assess the validity of his measure.

McClelland selected items for his test in terms of their capacity to discriminate Ss rated high and low on criterion scales considered relevant to this ability. Ss were rated for social intelligence, reaction sensitivity, ability to predict behavior, rigidity, and self-objectivity. A sixth

scale, RPA, was discarded when his Ss claimed not to have information on which to base ratings.

In the present study, the 30 students in E's UCLA course in psychological interviewing filled out the RPA test early in the semester. Students also met weekly in role-playing laboratory groups throughout the semester. In the last session each S rated himself and the three to five other members of his laboratory group on all six criterion scales. RPA test scores were separately correlated with both self-ratings and the average of the other lab group members for the sum of ratings on all six criterion scales, the five scales McClelland used, and the RPA scale alone. Contrary to expectation almost half the correlations, including the only significant one, were negative.

In two class sessions towards the end of the semester two male sophomore volunteers, J and M, were interviewed separately. J and M then rank ordered 16 adjectives descriptive of interpersonal personal behavior in terms of how they had just behaved in the interview, and their general social behavior. Immediately after observing the interview Ss were asked to predict how J and M ranked the adjectives. In addition Ss predicted J's and M's responses to a 100-item sentence completion test presented in a five-alternative multiple-choice format. These three cross-prediction tasks served as criteria against which to validate the RPA test. For both J and M on all three criterion tasks the correlations between RPA score and cross-predictions were all of zero order, and many were negative. It was concluded that the questionnaire is not a valid measure of RPA.

REFERENCE

- MCCLELLAND, W. A. A preliminary test of role-playing ability. *J. consult. Psychol.*, 1951, 15, 2, 102-108.

(Received December 8, 1959)

¹This study was conducted at UCLA, supported by a grant from the Faculty Research Committee. The contributions of Richard Timmer, Mark Golden, and David Raskin are gratefully acknowledged.

An extended report of this study may be obtained without charge from Zanwil Sperber (Research Department, Children's Hospital of Philadelphia; 1740 Bainbridge Street; Philadelphia 46, Pennsylvania) or for a fee from the American Documentation Institute. Order Document No. 6409 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

THE GENERALITY OF EXPECTANCY STATEMENTS AS A FUNCTION OF SITUATIONAL DEFINITION¹

SANFORD J. DEAN

Syracuse University

Studies of the generality of goal setting behavior (level of aspiration) have reported correlations ranging from almost zero to as high as .93. Since many of these studies have regarded goal setting as a nonsituational dispositional variable, the failure to find any consistent generality has frequently been attributed to the unreliability of the test. Another possibility is that there is less generality to behavior than has been assumed by theorists relying on broad dispositional concepts. It is possible to regard situational variables, not as sources of error which result in unreliable measurement, but as necessary variables to take into account in any prediction of behavior. The present study was concerned with the use of instructions as a means of defining the psychological situation to the subject (S). It was hypothesized that with the physical similarity of two tasks held constant, the correlation between expectancy statements on the two tasks would be significantly higher when the definitions of the tasks were similar.

Sixty-six college students, 27 males and 39 females, from an introductory psychology class

¹ This study was carried out under provisions of Project NR 150-087 supported by Contract Nonr 225 (01) between Stanford University and the Office of Naval Research. Permission is granted to the United States Government for reproduction, translation, publication, use, and disposal in whole or in part.

An extended report of this study may be obtained without charge from Sanford J. Dean (Psychology Department, Syracuse University; Syracuse, New York) or for a fee from the American Documentation Institute. Order Document No. 6410 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

served as Ss. Each S was given eight trials on each of two level of aspiration tasks and was required to predict his score prior to each trial. The first task required S to record as many words as he could think of beginning with each of eight letters orally presented by the experimenter (E). Following each trial a weighted score was reported to S based on a scale from 1 to 100. All Ss received the same predetermined scores. The second task consisted of sorting a deck of 100 playing cards into a four compartment box according to suit. The score was the number correctly sorted in 30 seconds. All Ss were again given the same scores by the E calling time when the appropriate number of sortings had been achieved. Scores on the second task were a constant three points higher by trial than those given on the first task. The instructions for Task I were the same for all Ss and were to the effect that they were to take a short-form intelligence test which seemed to correlate well with other intelligence tests and with academic achievement. The instructions for Task II varied depending on the group. Alternate Ss were told that the second task was another short-form intelligence test or that it was a motor-speed test which seemed to correlate well with things requiring speed and manual dexterity.

The product-moment correlation for Group I between mean expectancy statements Trials 2-8 on the two tasks was .88. The corresponding correlations for Group II was .62. Both correlations are significant at less than the .001 level and the difference between them is significant at less than the .002 level. Under conditions designed to maximize generality, with physical similarity and performance controlled, the generality of expectancy statements was found to vary with the psychological situation as defined by the instructions.

(Received January 14, 1960)

